Submitted to Bernoulli

Non-stationary phase of the MALA algorithm

JUAN KUNTZ[1], MICHELA OTTOBRE[2] and ANDREW M. STUART[3]

[1] *Mathematics Department, Imperial College London, 180 Queen's Gate, SW7 2AZ, London, UK*
E-mail: juan.kuntz08@imperial.ac.uk

[2] *Mathematics Department, Heriot-Watt University, Edinburgh, EH14 4AS , UK*
E-mail: michelaottobre@gmail.com

[3] *Department of Computing and Mathematical Sciences, California Institute of Technology, CA 91125, USA*
E-mail: astuart@caltech.edu

The Metropolis-Adjusted Langevin Algorithm (MALA) is a Markov Chain Monte Carlo method which creates a Markov chain reversible with respect to a given target distribution, π^N, with Lebesgue density on $\mathbb{R}^N$; it can hence be used to approximately sample the target distribution. When the dimension N is large a key question is to determine the computational cost of the algorithm as a function of N. One approach to this question, which we adopt here, is to derive diffusion limits for the algorithm. The family of target measures that we consider in this paper are, in general, in non-product form and are of interest in applied problems as they arise in Bayesian nonparametric statistics and in the study of conditioned diffusions. In particular, we work in the setting in which families of measures on spaces of increasing dimension are found by approximating a measure on an infinite dimensional Hilbert space which is defined by its density with respect to a Gaussian. Furthermore, we study the situation, which arises in practice, where the algorithm is started out of stationarity. We thereby significantly extend previous works which consider either only measures of product form, when the Markov chain is started out of stationarity, or measures defined via a density with respect to a Gaussian, when the Markov chain is started in stationarity. We prove that, in the non-stationary regime, the computational cost of the algorithm is of the order $N^{1/2}$ with dimension, as opposed to what is known to happen in the stationary regime, where the cost is of the order $N^{1/3}$.

Keywords: Markov Chain Monte Carlo, Metropolis-Adjusted Langevin Algorithm, diffusion limit, optimal scaling.

1. Introduction

Metropolis-Hastings algorithms are popular MCMC methods used to sample from a given probability measure, referred to as the target measure. The basic mechanism consists of employing a proposal transition density $q(x, y)$ in order to produce a reversible chain $\{x^k\}_{k=0}^{\infty}$ which has the target measure π as invariant distribution [Tie98]. At step k of the chain, a proposal move y^k is generated by using $q(x, y)$, i.e. $y^k \sim q(x^k, \cdot)$. Then such a move is accepted with probability $\alpha(x^k, y^k)$:

$$\alpha(x^k, y^k) = \min\left\{1, \frac{\pi(y^k)q(y^k, x^k)}{\pi(x^k)q(x^k, y^k)}\right\}. \tag{1.1}$$

The present paper aims at studying the computational cost of the MALA algorithm, when such an algorithm is in its non-stationary regime and the measure π is in non-product form. We will first introduce the class of target measures that we consider and then clarify the problem that is subject of this paper.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ be an infinite dimensional separable Hilbert space and consider the measure π on $\mathcal{H}$, defined as follows:

$$\frac{d\pi}{d\pi_0} \propto \exp(-\Psi), \qquad \pi_0 \overset{\mathcal{D}}{\sim} \mathcal{N}(0, \mathcal{C}). \tag{1.2}$$

That is, π is absolutely continuous with respect to a Gaussian measure π_0 with mean zero and covariance operator $\mathcal{C}$. Ψ is some real valued functional with domain $\tilde{\mathcal{H}} \subseteq \mathcal{H}$, $\Psi : \tilde{\mathcal{H}} \to \mathbb{R}$. Measures of the form (1.2) naturally arise in Bayesian nonparametric statistics and in the study of conditioned diffusions [Stu10, HSVW05]. In Section 2 we will give the precise definition of the space $\tilde{\mathcal{H}}$ and identify it with an appropriate

Sobolev-like subspace of $\mathcal{H}$ (denoted by $\mathcal{H}^s$ in Section 2).The covariance operator $\mathcal{C}$ is a positive, self-adjoint, trace class operator on $\mathcal{H}$, with eigenbasis $\{\lambda_j^2, \phi_j\}$:

$$\mathcal{C}\phi_j = \lambda_j^2 \phi_j, \quad \forall j \in \mathbb{N}, \tag{1.3}$$

and we assume that the set $\{\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$.

We will analyse the MALA algorithm designed to sample from the finite dimensional projections π^N of the measure (1.2) on the space

$$\mathcal{H} \supset X^N := \mathrm{span}\{\phi_j\}_{j=1}^N \tag{1.4}$$

spanned by the first N eigenvectors of the covariance operator. Notice that the space X^N is isomorphic to $\mathbb{R}^N$. To clarify this further, we need to introduce some notation. Given a point $x \in \mathcal{H}$, $\mathcal{P}^N(x)$ is the projection of x onto the space X^N; with slight abuse of notation, we will also denote $\Psi^N(x) := \Psi(\mathcal{P}^N(x))$ and $\mathcal{C}_N$ will be, effectively, an $N \times N$ diagonal matrix with i-th diagonal component equal to λ_i^2. More formally,

$$\Psi^N := \Psi \circ \mathcal{P}^N \quad \text{and} \quad \mathcal{C}_N := \mathcal{P}^N \circ \mathcal{C} \circ \mathcal{P}^N. \tag{1.5}$$

With this notation in place, our target measure is the measure π^N (on $X^N \cong \mathbb{R}^N$) defined as

$$\frac{d\pi^N}{d\pi_0^N}(x) = M_{\Psi^N} e^{-\Psi^N(x)}, \qquad \pi_0^N \sim \mathcal{N}(0, \mathcal{C}_N), \tag{1.6}$$

where M_{Ψ^N} is a normalization constant. Notice that the sequence of measures $\{\pi^N\}_{N\in\mathbb{N}}$ approximates the measure π (in particular, the sequence $\{\pi^N\}_{N\in\mathbb{N}}$ converges to π in the Hellinger metric, see [Stu10, Section 4] and references therein). In order to sample from the measure π^N in (1.6), we will consider the MALA algorithm with proposal

$$y^{k,N} = x^{k,N} + \delta \nabla \log \pi^N(x^{k,N}) + \sqrt{2\delta}\, \mathcal{C}_N^{1/2} \xi^{k,N}, \tag{1.7}$$

where

$$\xi^{k,N} = \sum_{i=1}^N \xi_i \phi_i, \quad \xi_i \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1) \text{ i.i.d},$$

and $\delta > 0$ is a positive parameter. Roughly speaking, for any fixed $N \in \mathbb{N}$, the MALA algorithm produces a (N-dimensional) Markov chain $\{x^{k,N}\}_k \subseteq X^N$ as follows: if the chain is in $x^{k,N}$ at step k, the algorithm proposes a move to $y^{k,N}$, defined in (1.7). The move is then accepted or rejected according to the acceptance probability defined in (1.1) (with $q(x^{k,N}, \cdot)$ the proposal kernel implied by (1.7), see (3.5)). A detailed description of the algorithm will be given in Section 3. For the time being it suffices to say that a crucial parameter to be appropriately chosen in order to optimize the performance of the algorithm is the proposal variance (or, informally, the 'jump size') δ appearing in (1.7). The choice of the proposal variance, and in particular the optimal scaling of δ with N, will be our main subject of study in this paper. To explain the issue in more detail, set $\delta = \ell N^{-\zeta}$, where $\ell > 0$ and $\zeta > 0$ are positive parameters to be chosen. The latter parameter, ζ, is the most relevant to our discussion, so we focus on describing how the performance of MALA is affected by the choice of ζ. As is well known [RGG97, RR98, JLM15, JLM14], if ζ is too small (so that δ is too big) then the proposed moves are too far away from the current state and the algorithm tends to reject them very often, therefore moving slowly (and this is more and more the case as the dimension N increases). On the other hand, if ζ is too big (so that the jump size is too small) then the algorithm will accept the proposed moves more frequently but, because all the moves are close to each other, the chain will anyway explore the state space slowly and inefficiently. It is therefore clear that one needs to find the optimal value of ζ that strikes the balance between these two opposing scenarios and this is what we mean when we refer to the "optimal" choice of the proposal variance.

When the MALA algorithm is initialised in stationarity (that is, $x^{0,N}$ is distributed according to π^N), the optimal choice of scaling for δ is known to be $\delta = \ell/N^{1/3}$ (see [PST12, RR98] and Section 1.1 for a more careful literature comparison). In the present paper we prove that, if the algorithm is started out of stationarity, then, in the non-stationary regime, the optimal choice of scaling is given by

$$\delta = \frac{\ell}{\sqrt{N}}.$$

We fix the above choice of δ throughout the paper, unless otherwise stated. We will make further comments on this point and on related literature in Section 1.1. We now come to explain the main result of the paper.

Using the proposal (1.7) we construct the Metropolis-Hastings chain $\{x^{k,N}\}_{k\in\mathbb{N}}$ and consider the continuous interpolant

$$x^{(N)}(t) = (N^{1/2}t - k)x^{k+1,N} + (k+1-N^{1/2}t)x^{k,N}, \quad t_k \le t < t_{k+1}, \text{ where } t_k = \frac{k}{N^{1/2}}. \tag{1.8}$$

The main result of this paper is the diffusion limit for the MALA algorithm, which we informally state here. The precise statement of such a result is given in Theorem 5.2 (and Section 5 contains heuristic arguments which explain how such a result is obtained). Below $C([0,T];\tilde{\mathcal{H}})$ denotes the space of $\tilde{\mathcal{H}}$-valued continuous functions on $[0,T]$, endowed with the uniform topology; α_ℓ, h_ℓ and b_ℓ are real valued functions, which we will define immediately after the statement, and $x_j^{k,N}$ denotes the j-th component of the vector $x^{k,N} \in X^N$ with respect to the basis $\{\phi_1,\dots,\phi_N\}$ (more details on this notation are given in Subsection 2.1.)

Main Result. *Let $\{x^{k,N}\}_{k\in\mathbb{N}}$ be the Metropolis-Hastings Markov chain to sample from π^N and constructed using the MALA proposal (1.7) (i.e. the chain (3.8)). Then, for any deterministic initial datum $x^{0,N} = \mathcal{P}^N(x^0)$, where x^0 is any vector in $\tilde{\mathcal{H}}$, the continuous interpolant $x^{(N)}$ defined in (1.8) converges weakly in $C([0,T];\tilde{\mathcal{H}})$ to the solution of the SDE*

$$dx(t) = -h_\ell(S(t))[x(t) + \mathcal{C}\nabla\Psi(x(t))]\,dt + \sqrt{2h_\ell(S(t))}\,dW(t), \quad x(0) = x^0, \tag{1.9}$$

where $S(t) \in \mathbb{R}_+ := \{s \in \mathbb{R} : s \ge 0\}$ solves the ODE

$$dS(t) = b_\ell(S(t))\,dt, \qquad S(0) := \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \frac{\left|x_j^{0,N}\right|^2}{\lambda_j^2}. \tag{1.10}$$

In the above the initial datum $S(0)$ is assumed to be finite and $W(t)$ is a $\tilde{\mathcal{H}}$-valued Brownian motion with covariance $\tilde{\mathcal{C}}$. [1]

The functions $\alpha_\ell, h_\ell, b_\ell : \mathbb{R} \to \mathbb{R}$ in the previous statement are defined as follows:

$$\alpha_\ell(s) = 1 \wedge e^{\ell^2(s-1)/2} \tag{1.11}$$

$$h_\ell(s) = \ell\alpha_\ell(s) \tag{1.12}$$

$$b_\ell(s) = 2\ell(1-s)\left(1 \wedge e^{\ell^2(s-1)/2}\right) = 2(1-s)h_\ell(s). \tag{1.13}$$

Remark 1.1. We make several remarks concerning the main result.

- Since the effective time-step implied by the interpolation (1.8) is $N^{-1/2}$, the main result implies that the optimal scaling for the proposal variance when the chain is in its non- stationary regime is $\delta \propto N^{-1/2}$. More comments on this fact can be found in Section 5.
- Notice that equation (1.10) evolves independently of equation (1.9). Once the MALA algorithm (3.8) is introduced and an initial state $x^0 \in \tilde{\mathcal{H}}$ is given such that $S(0)$ is finite, the real valued (double) sequence $S^{k,N}$,

$$S^{k,N} := \frac{1}{N}\sum_{i=1}^{N} \frac{\left|x_i^{k,N}\right|^2}{\lambda_i^2} \tag{1.14}$$

  started at $S_0^N := \frac{1}{N}\sum_{i=1}^N \frac{\left|x_i^{0,N}\right|^2}{\lambda_i^2}$ is well defined. For fixed N, $\{S^{k,N}\}_k$ is not, in general, a Markov process (however it is Markov if e.g. $\Psi = 0$). Consider the continuous interpolant $S^{(N)}(t)$ of the sequence

[1] The operator that here we denote generically by $\tilde{\mathcal{C}}$, to avoid getting in too much notation at this stage, will be more clearly defined in Section 2 and there denoted by $\mathcal{C}_s$. More precisely, as we will explain, $W(t)$ is a Brownian motion with covariance $\mathcal{C}_s$, see Section 2, in particular (2.5) and (2.4).

$S^{k,N}$, namely

$$S^{(N)}(t) = (N^{1/2}t - k)S^{k+1,N} + (k + 1 - N^{1/2}t)S^{k,N}, \quad t_k \leq t < t_{k+1},\ t_k = \frac{k}{\sqrt{N}}. \tag{1.15}$$

In Theorem 5.1 we prove that $S^{(N)}(t)$ converges in probability in $C([0,T];\mathbb{R})$ to the solution of the ODE (1.10) with initial condition $S_0 := \lim_{N\to\infty} S_0^N$. Once such a result is obtained, we can prove that $x^{(N)}(t)$ converges to $x(t)$. We want to stress that the convergence of $S^{(N)}(t)$ to $S(t)$ can be obtained independently of the convergence of $x^{(N)}(t)$ to $x(t)$.

- Let $S(t) : \mathbb{R} \to \mathbb{R}$ be the solution of the ODE (1.10). We will prove (see Theorem 4.1) that $S(t) \to 1$ as $t \to \infty$. With this in mind, notice that $h_\ell(1) = \ell$. Heuristically one can then argue that the asymptotic behaviour of the law of $x(t)$, the solution of (1.9), is described by the law of the following infinite dimensional SDE:

$$dz(t) = -\ell(z + \mathcal{C}\nabla\Psi(z)) + \sqrt{2\ell}dW(t). \tag{1.16}$$

It was proved in [HSVW05, HSV07] that (1.16) is ergodic with unique invariant measure given by (1.2). Our deduction concerning computational cost is made on the assumption that the law of (1.9) does indeed tend to the law of (1.16), although we will not prove this here as it would take us away from the main goal of the paper which is to establish the diffusion limit of the MALA algorithm.

1.1. Related Literature

In the present paper we consider target measures in non-product form, when the chain is started out of stationarity. When the target measure is in product form, a diffusion limit for the resulting Markov chain was studied in the seminal paper [RGG97]. The work [RGG97] is carried out under the following two assumptions: i) the chain is started in stationarity; ii) the target measure p (on $\mathbb{R}^N$) is of the form

$$p(x^N) = \Pi_{i=1}^N e^{-V(x_i^N)}, \quad x^N := (x_1^N, \ldots, x_N^N) \in \mathbb{R}^N. \tag{1.17}$$

In the above the potential $V : \mathbb{R} \to \mathbb{R}$ is such that the measure p is normalized to be a probability measure. Under such assumptions it was shown that the optimal scaling of the proposal variance is $\delta \sim N^{-1/3}$, leading to the conclusion that, in stationarity, $\mathcal{O}(N^{1/3})$ steps are required to explore the target distribution. In [CRR05] the same question was addressed in the case where the chain is started out of stationarity and p is the density of a standard i.i.d. Gaussian, i.e. $p \sim \mathcal{N}(0, I_N)$, where I_N is the N-dimensional identity matrix. For this Gaussian i.i.d. case the authors prove that the optimal scaling is given by $\delta = \ell/N^\zeta$ with $\zeta = 1/3$ if we start in stationarity and $\zeta = 1/2$ if we start out of stationarity. The intuition behind the choice of scaling that we make in this paper is indeed dictated by the results of [CRR05] and the diffusion limit that we prove for $S^{(N)}$ can be seen as a generalization of [CRR05, Lemma 4]. In this paper we show that the same holds also for the more general non-product target (1.6) (more remarks on this point will be made in Section 5.1). Recently the papers [JLM15, JLM14] made the significant extension of considering the product case (1.17) for quite general potentials V, again out of stationarity. In such works the authors prove that, in the non-stationary regime, the optimal scaling for the MALA proposal will depend, in general, on the potential V. Again recently, diffusion limits for MALA started in stationarity have also been considered for measures in non-product form in [PST12], using families of target measures found by approximating (1.2), as we consider in this paper; once again the conclusion is that, in stationarity, $O(N^{1/3})$ steps are required to explore the target distribution. In the present paper we combine the settings of [PST12] and [JLM15] and make a further significant extension of the analysis to consider measures in non-product form, when the chain is started out of stationarity, showing that the optimal scaling of the jump size is $\delta \propto N^{-1/2}$ in the transient regime.

We do not describe here in detail the relation between our results and the results of [JLM15, JLM14]. We just mention that in [JLM15] the diffusion limit for the MALA algorithm started out of stationarity and targeting measures of the form (1.17) is given by a non-linear equation of McKean-Vlasov type. This is in contrast with our diffusion limit, which is an infinite-dimensional SDE. The reason why this is the case is discussed in detail in [KOS16, Section 1.2]. The discussion in the latter paper is referred to the Random

Walk Metropolis algorithm, but it is conceptually analogous to what holds for the MALA algorithm and for this reason we do not spell it out here.

We mention for completeness that the non stationary case has also been considered in [PST14, OPPS16], for the pCN (preconditioned Crank-Nicolson) algorithm and for the SOL-HMC (Second Order Langevin - Hamiltonian Monte Carlo) scheme, respectively. These algorithms are well-defined in the infinite dimensional limit and hence do not require a scaling of the time-step which is inversely proportional to a power of the dimension.

1.2. Structure of the paper

The paper is organized as follows. In Section 2 we introduce the notation that we will use in the rest of the paper and the assumptions on the functional Ψ and on the covariance operator $\mathcal{C}$. In Section 3 we present in more detail the MALA algorithm. Section 4 contains the proof of existence and uniqueness of solutions for the limiting equations (1.9) and (1.10). With these preliminaries in place, we give, in Section 5, the formal statement of the main results of this paper, Theorem 5.1 and Theorem 5.2. In Section 5 we also provide heuristic arguments to explain how the main results are obtained. Such arguments are then made rigorous in Section 7, Section 8 and Section 9. In particular, Section 7 contains preliminary estimates and the analysis of the acceptance probability; Section 8 and Section 9 contain the proof of Theorem 5.1 and Theorem 5.2, respectively. The continuous mapping argument these proofs rely on is presented in Section 6. The reader who wants to understand how the result is derived, without getting in too many details, can skip the next three sections and move to Section 5.

2. Notation and Assumptions

In this section we detail the notation and the assumptions (Section 2.1 and Section 2.2, respectively) that we will use in the rest of the paper.

2.1. Notation

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ denote a real separable infinite dimensional Hilbert space, with the canonical norm induced by the inner-product. Let π_0 be a zero-mean Gaussian measure on $\mathcal{H}$ with covariance operator $\mathcal{C}$. By the general theory of Gaussian measures [DZ92], $\mathcal{C}$ is a positive, trace class operator. Let $\{\phi_j, \lambda_j^2\}_{j \geq 1}$ be the eigenfunctions and eigenvalues of $\mathcal{C}$, respectively, so that (1.3) holds. We assume a normalization under which $\{\phi_j\}_{j \geq 1}$ forms a complete orthonormal basis of $\mathcal{H}$. Recalling (1.4), we specify the notation that will be used throughout this paper:

- x and y are elements of the Hilbert space $\mathcal{H}$;
- the letter N is reserved to denote the dimensionality of the space X^N where the target measure π^N is supported;
- x^N is an element of $X^N \cong \mathbb{R}^N$ (similarly for y^N and the noise ξ^N);
- for any fixed $N \in \mathbb{N}$, $x^{k,N}$ is the k-th step of the chain $\{x^{k,N}\}_{k \in \mathbb{N}} \subseteq X^N$ constructed to sample from π^N; $x_i^{k,N}$ is the i-th component of the vector $x^{k,N}$, that is $x_i^{k,N} := \langle x^{k,N}, \phi_i \rangle$ (with abuse of notation).

For every $x \in \mathcal{H}$, we have the representation $x = \sum_{j \geq 1} x_j \phi_j$, where $x_j := \langle x, \phi_j \rangle$. Using this expansion, we define Sobolev-like spaces $\mathcal{H}^s$, $s \in \mathbb{R}$, with the inner-products and norms defined by

$$\langle x, y \rangle_s = \sum_{j=1}^{\infty} j^{2s} x_j y_j \qquad \text{and} \qquad \|x\|_s^2 = \sum_{j=1}^{\infty} j^{2s} x_j^2.$$

The space $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ is also a Hilbert space. Notice that $\mathcal{H}^0 = \mathcal{H}$. Furthermore $\mathcal{H}^s \subset \mathcal{H} \subset \mathcal{H}^{-s}$ for any $s > 0$. The Hilbert-Schmidt norm $\| \cdot \|_{\mathcal{C}}$ associated with the covariance operator $\mathcal{C}$ is defined as

$$||x||_{\mathcal{C}}^2 := \sum_{j=1}^{\infty} \lambda_j^{-2} x_j^2 = \sum_{j=1}^{\infty} \frac{|\langle x, \phi_j \rangle|^2}{\lambda_j^2}, \qquad x \in \mathcal{H},$$

and it is the Cameron-Martin norm associated with the Gaussian measure $\mathcal{N}(0,\mathcal{C})$. Such a norm is induced by the scalar product

$$\langle x,y\rangle_{\mathcal{C}} := \langle \mathcal{C}^{-1/2}x, \mathcal{C}^{-1/2}y\rangle, \qquad x,y\in\mathcal{H}.$$

Similarly, $\mathcal{C}_N$ defines a Hilbert-Schmidt norm on X^N,

$$||x^N||^2_{\mathcal{C}_N} := \sum_{j=1}^N \frac{\left|\langle x^N,\phi_j\rangle\right|^2}{\lambda_j^2}, \qquad x^N\in X^N, \tag{2.1}$$

which is induced by the scalar product

$$\langle x^N,y^N\rangle_{\mathcal{C}_N} := \langle \mathcal{C}^{-1/2}x^N, \mathcal{C}^{-1/2}y^N\rangle, \qquad x^N,y^N\in X^N.$$

For $s\in\mathbb{R}$, let $L_s:\mathcal{H}\to\mathcal{H}$ denote the operator which is diagonal in the basis $\{\phi_j\}_{j\geq 1}$ with diagonal entries j^{2s},

$$L_s\,\phi_j = j^{2s}\phi_j,$$

so that $L_s^{\frac{1}{2}}\,\phi_j = j^s\phi_j$. The operator L_s lets us alternate between the Hilbert space $\mathcal{H}$ and the interpolation spaces $\mathcal{H}^s$ via the identities:

$$\langle x,y\rangle_s = \langle L_s^{\frac{1}{2}}x, L_s^{\frac{1}{2}}y\rangle \qquad \text{and} \qquad \|x\|_s^2 = \|L_s^{\frac{1}{2}}x\|^2.$$

Since $\left|\left|L_s^{-1/2}\phi_k\right|\right|_s = ||\phi_k|| = 1$, we deduce that $\{\hat{\phi}_k := L_s^{-1/2}\phi_k\}_{k\geq 1}$ forms an orthonormal basis of $\mathcal{H}^s$. An element $y\sim\mathcal{N}(0,\mathcal{C})$ can be expressed as

$$y=\sum_{j=1}^\infty \lambda_j\rho_j\phi_j \qquad \text{with} \qquad \rho_j \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1) \text{ i.i.d.} \tag{2.2}$$

If $\sum_j \lambda_j^2 j^{2s}<\infty$, then y can be equivalently written as

$$y=\sum_{j=1}^\infty (\lambda_j j^s)\rho_j(L_s^{-1/2}\phi_j) \qquad \text{with} \qquad \rho_j \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1) \text{ i.i.d.} \tag{2.3}$$

For a positive, self-adjoint operator $D:\mathcal{H}\mapsto\mathcal{H}$, its trace in $\mathcal{H}$ is defined as

$$\mathrm{Trace}_{\mathcal{H}}(D) := \sum_{j=1}^\infty \langle\phi_j, D\phi_j\rangle.$$

We stress that in the above $\{\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis for $(\mathcal{H},\langle\cdot,\cdot\rangle)$. Therefore, if $\tilde{D}:\mathcal{H}^s\to\mathcal{H}^s$, its trace in $\mathcal{H}^s$ is

$$\mathrm{Trace}_{\mathcal{H}^s}(\tilde{D}) = \sum_{j=1}^\infty \langle L_s^{-\frac{1}{2}}\phi_j, \tilde{D}L_s^{-\frac{1}{2}}\phi_j\rangle_s.$$

Since $\mathrm{Trace}_{\mathcal{H}^s}(\tilde{D})$ does not depend on the orthonormal basis, the operator $\tilde{D}$ is said to be trace class in $\mathcal{H}^s$ if $\mathrm{Trace}_{\mathcal{H}^s}(\tilde{D})<\infty$ for some, and hence any, orthonormal basis of $\mathcal{H}^s$. Because $\mathcal{C}$ is defined on $\mathcal{H}$, the covariance operator

$$\mathcal{C}_s = L_s^{1/2}\mathcal{C}L_s^{1/2} \tag{2.4}$$

is defined on $\mathcal{H}^s$. Thus, for all the values of r such that $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)=\sum_j\lambda_j^2 j^{2s}<\infty$, we can think of y as a mean zero Gaussian random variable with covariance operator $\mathcal{C}$ in $\mathcal{H}$ and $\mathcal{C}_s$ in $\mathcal{H}^s$ (see (2.2) and (2.3)). In the same way, if $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)<\infty$, then

$$W(t)=\sum_{j=1}^\infty \lambda_j w_j(t)\phi_j = \sum_{j=1}^\infty \lambda_j j^r w_j(t)\hat{\phi}_j, \tag{2.5}$$

where $\{w_j(t)\}_{j\geq 1}$ a collection of i.i.d. standard Brownian motions on $\mathbb{R}$, can be equivalently understood as an $\mathcal{H}$-valued $\mathcal{C}$-Brownian motion or as an $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion.

We will make use of the following elementary inequality,

$$|\langle x, y\rangle|^2 = \left|\sum_{j=1}^{\infty}(j^s x_j)(j^{-s}y_j)\right|^2 \leq ||x||_s^2\,||y||_{-s}^2\,, \qquad \forall x\in\mathcal{H}^s, \quad y\in\mathcal{H}^{-s}\,. \tag{2.6}$$

Throughout this paper we study sequences of real numbers, random variables and functions, indexed by either (or both) the dimension N of the space on which the target measure is defined or the chain's step number k. In doing so, we find the following notation convenient.

- Two (double) sequences of real numbers $\{A^{k,N}\}$ and $\{B^{k,N}\}$ satisfy $A^{k,N} \lesssim B^{k,N}$ if there exists a constant $K>0$ (independent of N and k) such that

  $$A^{k,N}\leq KB^{k,N},$$

  for all N and k such that $\{A^{k,N}\}$ and $\{B^{k,N}\}$ are defined.
- If the $A^{k,N}$s and $B^{k,N}$s are random variables, the above inequality must hold almost surely (for some deterministic constant K).
- If the $A^{k,N}$s and $B^{k,N}$s are real-valued functions on $\mathcal{H}$ or $\mathcal{H}^s$, $A^{k,N}=A^{k,N}(x)$ and $B^{k,N}=B^{k,N}(x)$, the same inequality must hold with K independent of x, for all x where the $A^{k,N}$s and $B^{k,N}$s are defined.

As customary, $\mathbb{R}_+ := \{s\in\mathbb{R} : s\geq 0\}$ and for all $b\in\mathbb{R}_+$ we let $[b]=n$ if $n\leq b<n+1$ for some integer n. Finally, for time dependent functions we will use both the notations $S(t)$ and S_t interchangeably.

2.2. Assumptions

In this section we describe the assumptions on the covariance operator $\mathcal{C}$ of the Gaussian measure $\pi_0 \overset{\mathcal{D}}{\sim} \mathcal{N}(0,\mathcal{C})$ and those on the functional Ψ. We fix a distinguished exponent $s\geq 0$ and assume that $\Psi:\mathcal{H}^s\to\mathbb{R}$ and $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)<\infty$. In other words $\mathcal{H}^s$ is the space that we were denoting with $\tilde{\mathcal{H}}$ in the introduction. Since

$$\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)=\sum_{j=1}^{\infty}\lambda_j^2 j^{2s}, \tag{2.7}$$

the condition $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)<\infty$ implies that $\lambda_j j^s\to 0$ as $j\to\infty$. Therefore the sequence $\{\lambda_j j^s\}_j$ is bounded:

$$\lambda_j j^s \leq C, \tag{2.8}$$

for some constant $C>0$ independent of j.

For each $x\in\mathcal{H}^s$ the derivative $\nabla\Psi(x)$ is an element of the dual $\mathcal{L}(\mathcal{H}^s,\mathbb{R})$ of $\mathcal{H}^s$, comprising the linear functionals on $\mathcal{H}^s$. However, we may identify $\mathcal{L}(\mathcal{H}^s,\mathbb{R})=\mathcal{H}^{-s}$ and view $\nabla\Psi(x)$ as an element of $\mathcal{H}^{-s}$ for each $x\in\mathcal{H}^s$. With this identification, the following identity holds

$$||\nabla\Psi(x)||_{\mathcal{L}(\mathcal{H}^s,\mathbb{R})} = ||\nabla\Psi(x)||_{-s}\,.$$

To avoid technicalities we assume that the gradient of $\Psi(x)$ is bounded and globally Lipschitz. More precisely, throughout this paper we make the following assumptions.

Assumptions 2.1. *The functional Ψ and covariance operator $\mathcal{C}$ satisfy the following:*

1. **Decay of Eigenvalues λ_j^2 of $\mathcal{C}$:** *there exists a constant $\kappa>\frac{1}{2}$ such that*

$$\lambda_j \asymp j^{-\kappa}.$$

2. **Domain of Ψ:** *there exists an exponent $s \in [0, \kappa - 1/2)$ such that Ψ is defined everywhere on $\mathcal{H}^s$.*
3. **Derivatives of Ψ:** *The derivative of Ψ is bounded and globally Lipschitz:*

$$||\nabla\Psi(x)||_{-s} \lesssim 1, \qquad ||\nabla\Psi(x) - \nabla\Psi(y)||_{-s} \lesssim ||x - y||_s . \tag{2.9}$$

Remark 2.2. The condition $\kappa > \frac{1}{2}$ ensures that $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$ for any $0 \le s < \kappa - \frac{1}{2}$. Consequently, π_0 has support in $\mathcal{H}^s$ ($\pi_0(\mathcal{H}^s) = 1$) for any $0 \le s < \kappa - \frac{1}{2}$. $\square$

Example 2.3. The functional $\Psi(x) = \sqrt{1 + ||x||_s^2}$ satisfies all of the above. $\square$

Remark 2.4. Our assumptions on the change of measure (that is, on Ψ) are less general than those adopted in [KOS16, PST12] and related literature (see references therein). This is for purely technical reasons. In this paper we assume that Ψ grows linearly. If Ψ was assumed to grow quadratically, which is the case in the mentioned works, finding bounds on the moments of the chain $\{x^{k,N}\}_{k\ge1}$ (much needed in all of the analysis) would become more involved than it already is, see Remark B.1. However, under our assumptions, the measure π (or π^N) is still, in general, a fully non-product measure. $\square$

We now explore the consequences of Assumptions 2.1.

Lemma 2.5. *Let Assumptions 2.1 hold. Then*

1. *The function $\mathcal{C}\nabla\Psi(x)$ is bounded and globally Lipschitz on $\mathcal{H}^s$, that is*

$$||\mathcal{C}\nabla\Psi(x)||_s \lesssim 1 \quad \textit{and} \quad ||\mathcal{C}\nabla\Psi(x) - \mathcal{C}\nabla\Psi(y)||_s \lesssim ||x - y||_s . \tag{2.10}$$

Therefore, the function $F(z) := -z - \mathcal{C}\nabla\Psi(z)$ satisfies

$$||F(x) - F(y)||_s \lesssim ||x - y||_s \quad \textit{and} \quad ||F(x)||_s \lesssim 1 + ||x||_s . \tag{2.11}$$

2. *The function $\Psi(x)$ is globally Lipschitz and therefore also $\Psi^N(x) := \Psi(\mathcal{P}^N(x))$ is globally Lipschitz:*

$$\left|\Psi^N(y) - \Psi^N(x)\right| \lesssim ||y - x||_s . \tag{2.12}$$

Proof. The bounds (2.10) are a consequence of (2.9). We show how to obtain the second bound in (2.10):

$$\begin{aligned}
||\mathcal{C}\nabla\Psi(x) - \mathcal{C}\nabla\Psi(y)||_s^2 &= \sum_{j=1}^{\infty} \lambda_j^4 j^{2s} \left[(\nabla\Psi(x) - \nabla\Psi(y))_j\right]^2 \\
&= \sum_{j=1}^{\infty} (\lambda_j j^s)^4 j^{-2s} \left[(\nabla\Psi(x) - \nabla\Psi(y))_j\right]^2 \\
&\lesssim \|\nabla\Psi(x) - \nabla\Psi(y)\|_{-s}^2 \overset{(2.9)}{\lesssim} \|x - y\|_s^2,
\end{aligned}$$

where in the above we have used (2.8) and $(\nabla\Psi(x) - \nabla\Psi(y))_j$ denotes the j-th component of the vector $\nabla\Psi(x) - \nabla\Psi(y)$. With analogous calculations one can obtain the first bound in (2.10). As for the second equation in (2.11):

$$||F(z)||_s \lesssim ||z||_s + \|\mathcal{C}\nabla\Psi(z)\|_s \overset{(2.10)}{\lesssim} 1 + ||z||_s .$$

Similarly for the first bound in (2.11). The proof of equation (2.12) is standard, so we only sketch it: consider a line joining points x and y, $\gamma(t) = x + t(y - x), t \in [0, 1]$. Then

$$\begin{aligned}
\Psi(\gamma(1)) - \Psi(\gamma(0)) &= \Psi(y) - \Psi(x) \\
&= \int_0^1 dt \, \langle \nabla\Psi(\gamma(t)), y - x\rangle \lesssim ||y - x||_s ,
\end{aligned}$$

having used (2.9) and (2.6) in the last inequality. An analogous calculation to the above can be done for Ψ^N, after proving (2.14) below. $\square$

Before stating the next lemma, we observe that by definition of the projection operator $\mathcal{P}^N$ we have that

$$\nabla\Psi^N = \mathcal{P}^N \circ \nabla\Psi \circ \mathcal{P}^N. \tag{2.13}$$

Lemma 2.6. *Let Assumptions 2.1 hold. Then the following holds for the function Ψ^N and for its the gradient:*

1. *If the bounds (2.9) hold for Ψ, then they hold for Ψ^N as well:*

$$\left|\left|\nabla\Psi^N(x)\right|\right|_{-s} \lesssim 1, \qquad \left|\left|\nabla\Psi^N(x) - \nabla\Psi^N(y)\right|\right|_{-s} \lesssim \|x-y\|_s \,. \tag{2.14}$$

2. *Moreover,*

$$\left|\left|\mathcal{C}_N\nabla\Psi^N(x)\right|\right|_s \lesssim 1, \tag{2.15}$$

and

$$\left|\left|\mathcal{C}_N\nabla\Psi^N(x)\right|\right|_{\mathcal{C}_N} \lesssim 1. \tag{2.16}$$

We stress that in (2.14)-(2.16) the constant implied by the use of the notation "$\lesssim$" (see end of Section 2.1) is independent of N.

Proof of Lemma 2.6. The bounds (2.14) and (2.15) are just consequences of the definition of Ψ^N and $\nabla\Psi^N$ and the analogous properties of Ψ. For the sake of clarity we just spell out how to obtain (2.15):

$$\begin{aligned}\left|\left|\mathcal{C}_N\nabla\Psi^N(x)\right|\right|_s^2 &\overset{(2.13)}{=} \left|\left|\mathcal{C}_N\mathcal{P}^N\nabla\Psi(\mathcal{P}^N(x))\right|\right|_s^2 = \sum_{j=1}^N j^{2s}\lambda_j^4\left[\nabla\Psi(\mathcal{P}^N(x))\right]_j^2 \\ &\le \sum_{j=1}^\infty j^{2s}\lambda_j^4\left[\nabla\Psi(\mathcal{P}^N(x))\right]_j^2 \le \left|\left|\mathcal{C}\nabla\Psi(\mathcal{P}^N(x))\right|\right|_s^2 \overset{(2.10)}{\lesssim} 1\,.\end{aligned}$$

As for (2.16), using (2.8):

$$\|\mathcal{C}_N\nabla\Psi^N(x)\|_{\mathcal{C}_N}^2 = \sum_{j=1}^N \lambda_j^2\left[\left(\nabla\Psi^N(x)\right)_j\right]^2 \lesssim \sum_{j=1}^\infty j^{-2s}\left[\left(\nabla\Psi^N(x)\right)_j\right]^2 = \|\nabla\Psi^N(x)\|_{-s}^2 \lesssim 1.$$

□

We would also like to recall that because of our assumptions on the covariance operator,

$$\mathbb{E}\left|\left|\mathcal{C}_N^{1/2}\xi^N\right|\right|_s^2 \lesssim 1, \quad \text{uniformly in } N, \tag{2.17}$$

where $\xi^N := \sum_{j=1}^N \xi_j\phi_j$ and $\xi_i \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1)$ i.d.d., see [MPS12, (2.32)] or [KOS16, first proof of Appendix A]

3. The algorithm

The MALA algorithm stems from the observation that π^N is the unique stationary measure of the SDE

$$dY_t = \nabla\log\pi^N(Y_t)dt + \sqrt{2}dW_t^N, \tag{3.1}$$

where W^N is an X^N-valued Brownian motion with covariance operator $\mathcal{C}_N$. The algorithm consists of discretising (3.1) using the Euler-Maruyama scheme and adding a Metropolis accept-reject step so that the invariance of π^N is preserved. The MALA algorithm to sample from π^N is therefore a Metropolis-Hastings algorithm with proposal

$$y^{k,N} = x^{k,N} - \delta\left(x^{k,N} + \mathcal{C}_N\nabla\Psi^N(x^{k,N})\right) + \sqrt{2\delta}\mathcal{C}_N^{1/2}\xi^{k,N},^{2} \tag{3.2}$$

[2] In this paper the proposal move from step k is denoted by $y^{k,N}$. In [] it is denoted by y^{k+1}. We flag this up as the two papers naturally compare. Same observation applies to ξ^k and γ^k.

where

$$\xi^{k,N} := \sum_{j=1}^{N} \xi_j^{k,N} \phi_j, \quad \xi_j^{k,N} \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

We stress that the Gaussian random variables $\xi_i^{k,N}$ are independent of each other and of the current position $x^{k,N}$. Motivated by the considerations made in the introduction (and that will be made more explicit in Section 5.1), in this paper we fix the choice

$$\delta := \frac{\ell}{N^{1/2}}. \tag{3.3}$$

If at step k the chain is at $x^{k,N}$, the algorithm proposes a move to $y^{k,N}$ defined by equation (3.2). The move is then accepted with probability

$$\alpha^N(x^{k,N}, y^{k,N}) := \frac{\pi^N(y^{k,N}) q^N(y^{k,N}, x^{k,N})}{\pi^N(x^{k,N}) q^N(x^{k,N}, y^{k,N})}, \tag{3.4}$$

where, for any $x^N, y^N \in \mathbb{R}^N \simeq X^N$,

$$q^N(x^N, y^N) \propto e^{-\frac{1}{4\delta}\|(y^N - x^N) - \delta \nabla \log \pi^N(x^N)\|^2_{\mathcal{C}_N}}. \tag{3.5}$$

If the move to $y^{k,N}$ is accepted then $x^{k+1,N} = y^{k,N}$, if it is rejected the chain remains where it was, i.e. $x^{k+1,N} = x^{k,N}$. In short, the MALA chain is defined as follows:

$$x^{k+1,N} := \gamma^{k,N} y^{k,N} + (1 - \gamma^{k,N}) x^{k,N}, \qquad x^{0,N} := \mathcal{P}^N(x^0) \tag{3.6}$$

where in the above

$$\gamma^{k,N} \overset{\mathcal{D}}{\sim} \text{Bernoulli}(\alpha^N(x^{k,N}, y^{k,N})); \tag{3.7}$$

that is, conditioned on $(x^{k,N}, y^{k,N})$, $\gamma^{k,N}$ has Bernoulli law with mean $\alpha^N(x^{k,N}, y^{k,N})$. Equivalently, we can write

$$\gamma^{k,N} = \mathbf{1}_{\{U^{k,N} \le \alpha^N(x^{k,N}, y^{k,N})\}},$$

with $U^{k,N} \overset{\mathcal{D}}{\sim}$ Uniform $[0,1]$, independent of $x^{k,N}$ and $\xi^{k,N}$.

For fixed N, the chain $\{x^{k,N}\}_{k \ge 1}$ lives in $X^N \cong \mathbb{R}^N$ and samples from π^N. However, in view of the fact that we want to study the scaling limit of such a chain as $N \to \infty$, the analysis is cleaner if it is carried out in $\mathcal{H}$; therefore, the chain that we analyse is the chain $\{x^k\}_k \subseteq \mathcal{H}$ defined as follows: the first N components of the vector $x^k \in \mathcal{H}$ coincide with $x^{k,N}$ as defined above; the remaining components are not updated and remain equal to their initial value. More precisely, using (3.2) and (3.6), the chain x^k can be written in a component-wise notation as follows:

$$x_i^{k+1} = x_i^{k+1,N} = x_i^{k,N} - \gamma^{k,N} \left[\frac{\ell}{N^{1/2}} \left(x_i^{k,N} + [\mathcal{C}_N \nabla \Psi^N(x^{k,N})]_i \right) + \sqrt{\frac{2\ell}{N^{1/2}}} \lambda_i \, \xi^{k,N} \right] \tag{3.8}$$

for $i = 1, \dots, N$, while

$$x^{k+1} = x^k = x^0 \qquad \text{on } \mathcal{H} \setminus X^N.$$

For the sake of clarity, we specify that $[\mathcal{C}_N \nabla \Psi^N(x^{k,N})]_i$ denotes the i-th component of the vector $\mathcal{C}_N \nabla \Psi^N(x^{k,N}) \in \mathcal{H}^s$. From the above it is clear that the update rule (3.8) only updates the first N coordinates (with respect to the eigenbasis of $\mathcal{C}$) of the vector x^k. Therefore the algorithm evolves in the finite-dimensional subspace X^N. From now on we will avoid using the notation $\{x^k\}_k$ for the "extended chain" defined in $\mathcal{H}$, as it can be confused with the notation x^N, which instead is used throughout to denote a generic element of the space X^N.

We conclude this section by remarking that, if $x^{k,N}$ is given, the proposal $y^{k,N}$ only depends on the Gaussian noise $\xi^{k,N}$. Therefore the acceptance probability will be interchangeably denoted by $\alpha^N(x^N, y^N)$ or $\alpha^N(x^N, \xi^N)$.

4. Existence and uniqueness for limiting infinite dimensional SDE

The main results of this section are Theorem 4.1, Theorem 4.3 and Theorem 4.5. Theorem 4.1 and Theorem 4.3 are concerned with establishing existence and uniqueness for equations (1.9) and (1.10), respectively. Theorem 4.5 states the continuity of the Itô maps associated with equations (1.9) and (1.10). The proofs of the main results of this paper (Theorem 5.1 and Theorem 5.2) rely heavily on the continuity of such maps, as we illustrate in Section 6.

The proofs of the results of this section are completely analogous to the proofs of the results of [KOS16, Section 4]. We therefore only sketch them and refer the reader to [KOS16] for more details.

Theorem 4.1. *For any initial datum $S(0) \in \mathbb{R}_+$, there exists a unique solution $S(t) \in \mathbb{R}$ to the ODE (1.10). Such a solution is strictly positive for any $t > 0$, it is bounded and has continuous first derivative for all $t \geq 0$. In particular*

$$\lim_{t\to\infty} S(t) = 1$$

and

$$0 \leq \min\{S(0), 1\} \leq S(t) \leq \max\{S(0), 1\}\,. \tag{4.1}$$

Proof. Once the statement of Lemma 4.2 below is proved, the proof of the above theorem is completely analogous to the proof of [KOS16, Theorem 4.1]. □

We recall that the definition of the functions α_ℓ, h_ℓ and b_ℓ has been given in (1.11), (1.12) and (1.13), respectively.

Lemma 4.2. *The functions $\alpha_\ell(s)$, $h_\ell(s)$ and $\sqrt{h_\ell(s)}$ are positive, globally Lipschitz continuous and bounded. The function $b_\ell(s)$ is globally Lipschitz and it is bounded above but not below. Moreover, for any $\ell > 0$, $b_\ell(s)$ is strictly positive for $s \in [0, 1)$, strictly negative for $s > 1$ and $b_\ell(1) = 0$.*

Proof of Lemma 4.2. When $s > 1$, $\alpha_\ell(s) = 1$ while for $s \leq 1$ $\alpha_\ell(s)$ has bounded derivative; therefore $\alpha_\ell(s)$ is globally Lipshitz. A similar reasoning gives the Lipshitzianity of the other functions. The further properties of b_ℓ are straightforward from the definition. □

We now come to existence and uniqueness for equation (1.9), which we rewrite using the notation of Lemma 2.5 as

$$dx(t) = -h_\ell(S(t))F(x(t))\,dt + \sqrt{2h_\ell(S(t))}\,dW(t),$$

where $W(t)$ is an $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion. The above is intended to mean

$$x(t) = x(0) + \int_0^t F(x(v))h_\ell(S(v))dv + \int_0^t \sqrt{2h_\ell(S(v))}dW(v)\,. \tag{4.2}$$

Theorem 4.3. *Let Assumption 2.1 hold and consider equation (1.9) (or, equivalently, equation (4.2)), where $W(t)$ is any $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion and $S(t)$ is the solution of (1.10). Then for any initial condition $x(0) \in \mathcal{H}^s$ and any $T > 0$ there exists a unique solution of equation (1.9) in the space $C([0, T]; \mathcal{H}^s)$.*

Proof. With the statement of Theorem 4.1 and Lemma 4.2 in place, the proof is completely analogous to the proof of [KOS16, Theorem 4.3], so we omit it here. □

Consider now the following equation:

$$dx(t) = [-x(t) - \mathcal{C}\nabla\Psi(x(t))]h_\ell(S(t))\,dt + d\zeta(t), \tag{4.3}$$

where $S(t)$ is the solution of (1.10) and $\zeta(t)$ is any function in $C([0, T]; \mathcal{H}^s)$. Also, let $\mathfrak{S}(t) : \mathbb{R}_+ \to \mathbb{R}$ be the solution of

$$d\mathfrak{S}(t) = b_\ell(\mathfrak{S}(t))\,dt + a\,dw(t), \tag{4.4}$$

where $w(t)$ is a real valued standard Brownian motion and $a \in \mathbb{R}_+$ is a constant. Also, throughout the paper the spaces $C([0, T]; \mathcal{H}^s)$ and $C([0, T]; \mathbb{R})$ are assumed to be endowed with the uniform topology.

Remarks 4.4. Before stating the next theorem we need to be more precise about equations (4.3) and (4.4).

- We consider equation (4.4) (which is (1.10) perturbed by noise) in view of the contraction mapping argument (explained in Section 6) that we will use to prove our main results. Observe that (4.4) admits a unique solution, thanks to the Lipschitzianity of b_ℓ. Existence and uniqueness of the solution of (4.3) can be done with identical arguments to those used to prove existence and uniqueness of the solution to (1.9).
- We emphasize that (4.3) and (4.4) are decoupled as the function $S(t)$ appearing in (4.3) is the solution of (1.10). This fact will be particularly relevant in the remainder of this section as well as in Section 6.1 and Section 6.2.

□

The statement of the following theorem is crucial to the proof of our main result.

Theorem 4.5. *With the notation introduced so far (and in particular with the clarifications of Remarks 4.4) let $x(t)$ and $\mathfrak{S}(t)$ be the solutions of (4.3) and (4.4), respectively. Then, under Assumption 2.1, the Itô maps*

$$\begin{gathered}\mathcal{J}_1 : \mathcal{H}^s \times C([0,T];\mathcal{H}^s) \longrightarrow C([0,T];\mathcal{H}^s \times \mathbb{R})\\ (x_0, \zeta(t)) \longrightarrow x(t)\end{gathered}$$

and

$$\begin{gathered}\mathcal{J}_2 : \mathbb{R}_+ \times C([0,T];\mathbb{R}) \longrightarrow C([0,T];\mathbb{R})\\ (\mathfrak{S}_0, w(t)) \longrightarrow \mathfrak{S}(t)\end{gathered}$$

are continuous maps.

Proof. Analogous to the proof of [KOS16, Theorem 4.6]. □

5. Statement of main theorems and Heuristics of proofs

In order to state the main results, we first set

$$\mathcal{H}^s_\cap := \left\{ x \in \mathcal{H}^s : \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{|x_i|^2}{\lambda_i^2} < \infty \right\}, \tag{5.1}$$

where we recall that in the above $x_i := \langle x, \phi_i \rangle$.

Theorem 5.1. *Let Assumption 2.1 hold. Let $x^0 \in \mathcal{H}^s_\cap$ and $T > 0$. Then, as $N \to \infty$, the continuous interpolant $S^{(N)}(t)$ of the sequence $\{S^{k,N}\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_+$ (defined in (1.15)) and started at $S^{0,N} = \frac{1}{N}\sum_{i=1}^N \left|x_i^0\right|^2/\lambda_i^2$, converges in probability in $C([0,T];\mathbb{R})$ to the solution $S(t)$ of the ODE (1.10) with initial datum $S^0 := \lim_{N\to\infty} S^{0,N}$.*

Theorem 5.2. *Let Assumption 2.1 hold. Let $x^0 \in \mathcal{H}^s_\cap$ and $T > 0$. Then, as $N \to \infty$, the continuous interpolant $x^{(N)}(t)$ of the chain $\{x^{k,N}\}_{k\in\mathbb{N}} \subseteq \mathcal{H}^s$ (defined in (1.8) and (3.8), respectively) with initial state $x^{0,N} := \mathcal{P}^N(x^0)$, converges weakly in $C([0,T];\mathcal{H}^s)$ to the solution $x(t)$ of equation (1.9) with initial datum x^0. We recall that the time-dependent function $S(t)$ appearing in (1.9) is the solution of the ODE (1.10), started at $S(0) := \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \left|x_i^0\right|^2/\lambda_i^2$.*

Both Theorem 5.1 and Theorem 5.2 assume that the initial datum of the chains $x^{k,N}$ is assigned deterministically. From our proofs it will be clear that the same statements also hold for random initial data, as long as i) $x^{0,N}$ is not drawn at random from the target measure π^N or from any other measure which is a change of measure from π^N (i.e. we need to be starting out of stationarity) and ii) $S^{0,N}$ and $x^{0,N}$ have bounded moments (bounded uniformly in N) of sufficiently high order and are independent of all the other sources of noise present in the algorithm. Notice moreover that the convergence in probability of Theorem 5.1 is equivalent to weak convergence, as the limit is deterministic.

The rigorous proof of the above results is contained in Sections 6 to 9. In the remainder of this section we give heuristic arguments to justify our choice of scaling $\delta \propto N^{-1/2}$ and we explain how one can formally obtain the (fluid) ODE limit (1.10) for the double sequence $S^{k,N}$ and the diffusion limit (1.9) for the chain $x^{k,N}$. We stress that the arguments of this section are only formal; therefore, we often use the notation " $\simeq$ ", to mean "approximately equal". That is, we write $A \simeq B$ when $A = B+$ "terms that are negligible" as N tends to infinity; we then justify these approximations, and the resulting limit theorems, in the following Sections 6 to 9.

5.1. Heuristic analysis of the acceptance probability

As observed in [PST12, equation (2.21)], the acceptance probability (3.4) can be expressed as

$$\alpha^N(x^N, \xi^N) = 1 \wedge e^{Q^N(x^N, \xi^N)}, \tag{5.2}$$

where, using the notation (2.1), the function $Q^N(x, \xi)$ can be written as

$$\begin{aligned} Q^N(x^N, \xi^N) &:= -\frac{\delta}{4}\left(\left|\left|y^N\right|\right|^2_{\mathcal{C}_N} - \left|\left|x^N\right|\right|^2_{\mathcal{C}_N}\right) + r^N(x^N, \xi^N) \qquad (5.3)\\ &= \left[\frac{\delta^2}{2}\left(\left|\left|x^N\right|\right|^2_{\mathcal{C}_N} - \left|\left|\mathcal{C}_N^{1/2}\xi^N\right|\right|^2_{\mathcal{C}_N}\right)\right] - \frac{\delta^3}{4}\left|\left|x^N\right|\right|^2_{\mathcal{C}_N} \\ &\quad - \left(\frac{\delta^{3/2}}{\sqrt{2}} - \frac{\delta^{5/2}}{\sqrt{2}}\right)\langle x^N, \mathcal{C}_N^{1/2}\xi^N\rangle_{\mathcal{C}_N} + r^N_\Psi(x^N, \xi^N)\,. \qquad (5.4) \end{aligned}$$

We do not give here a complete expression for the terms $r^N(x^N, \xi^N)$ and $r^N_\Psi(x^N, \xi^N)$. For the time being it is sufficient to point out that

$$\begin{aligned} r^N(x^N, \xi^N) &:= I_2^N + I_3^N \\ r^N_\Psi(x^N, \xi^N) &:= r^N(x^N, \xi^N) + \frac{(\delta^2 - \delta^3)}{2}\langle x^N, \mathcal{C}_N \nabla\Psi^N(x^N)\rangle_{\mathcal{C}_N} \\ &\quad - \frac{\delta^3}{4}\|\mathcal{C}_N\nabla\Psi^N(x^N)\|^2_{\mathcal{C}_N} + \frac{\delta^{5/2}}{\sqrt{2}}\langle \mathcal{C}_N\nabla\Psi^N(x^N), \mathcal{C}_N^{1/2}\xi^N\rangle_{\mathcal{C}_N} \qquad (5.5) \end{aligned}$$

where I_2^N and I_3^N will be defined in (7.10) and (7.11), respectively. Because I_2^N and I_3^N depend on Ψ, r^N_Ψ contains all the terms where the functional Ψ appears; moreover r^N_Ψ vanishes when $\Psi = 0$. The analysis of Section 7 (see Lemma 7.5) will show that with our choice of scaling, $\delta = \ell/N^{1/2}$, the terms r^N and r^N_Ψ are negligible (for N large). Let us now illustrate the reason behind our choice of scaling. To this end, set $\delta = \ell/N^\zeta$ and observe the following two simple facts:

$$S^{k,N} = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|x_j^{k,N}\right|^2}{\lambda_j^2} = \frac{1}{N}\left|\left|x^{k,N}\right|\right|^2_{\mathcal{C}_N} \tag{5.6}$$

and

$$\left|\left|\mathcal{C}_N^{1/2}\xi^N\right|\right|^2_{\mathcal{C}_N} = \sum_{i=1}^{N}\left|\xi_i\right|^2 \simeq N, \tag{5.7}$$

the latter fact being true by the Law of Large Numbers. Neglecting the terms containing Ψ, at step k of the chain we have, formally,

$$Q^N(x^{k,N}, \xi^{k+1,N}) \simeq \frac{\ell^2}{2} N^{1-2\zeta} \left(S^{k,N} - 1\right) \tag{5.8}$$

$$- \frac{\ell^3}{4} N^{1-3\zeta} S^{k,N} - \frac{\ell^{3/2}}{\sqrt{2}} N^{(1-3\zeta)/2} \frac{\langle x^{k,N}, \mathcal{C}_N^{1/2} \xi^{k,N} \rangle_{\mathcal{C}_N}}{\sqrt{N}} \tag{5.9}$$

$$- \frac{\ell^{5/2}}{\sqrt{2}} N^{(1-5\zeta)/2} \frac{\langle x^{k,N}, \mathcal{C}_N^{1/2} \xi^{k,N} \rangle_{\mathcal{C}_N}}{\sqrt{N}}. \tag{5.10}$$

The above approximation (which, we stress again, is only formal and will be made rigorous in subsequent sections) has been obtained from (5.4) by setting $\delta = \ell / N^{\zeta}$ and using (5.6) and (5.7), as follows:

$$\frac{\delta^2}{2} \left[\left|\left| x^N \right|\right|^2_{\mathcal{C}_N} - \left|\left| \mathcal{C}_N^{1/2} \xi^N \right|\right|^2_{\mathcal{C}_N} \right] \simeq (5.8), \tag{5.11}$$

$$-\delta^3 \frac{\left|\left| x^N \right|\right|^2_{\mathcal{C}_N}}{4} - \frac{\delta^{3/2}}{\sqrt{2}} \langle x^N, \mathcal{C}_N^{1/2} \xi^N \rangle_{\mathcal{C}^N} \simeq (5.9),$$

$$-\frac{\delta^{5/2}}{\sqrt{2}} \langle x^N, \mathcal{C}_N^{1/2} \xi^N \rangle_{\mathcal{C}^N} = (5.10)\,.$$

Looking at the decomposition (5.8)-(5.10) of the function Q^N, we can now heuristically explain the reason why we are lead to choose $\zeta = 1/2$ when we start the chain out of stationarity, as opposed to the scaling $\zeta = 1/3$ when the chain is started in stationarity. This is explained in the following remarks.

Remarks 5.3. First notice that the expression (5.4) and the approximation (5.8)-(5.10) for Q^N are valid both in and out of stationarity, as the first is only a consequence of the definition of the Metropolis-Hastings algorithm and the latter is implied just by the properties of Ψ and by our definitions.

- If we start the chain in stationarity, i.e. $x_0^N \sim \pi^N$ (where π^N has been defined in (1.6)), then $x^{k,N} \sim \pi^N$ for every $k \geq 0$. As we have already observed, π^N is absolutely continuous with respect to the Gaussian measure $\pi_0^N \sim \mathcal{N}(0, \mathcal{C}_N)$; because all the almost sure properties are preserved under this change of measure, in the stationary regime most of the estimates of interest need to be shown only for $x^N \sim \pi_0^N$. In particular if $x^N \sim \pi_0^N$ then x^N can be represented as $x^N = \sum_{i=1}^N \lambda_i \rho_i \phi_i$, where ρ_i are i.i.d. $\mathcal{N}(0,1)$. Therefore we can use the law of large numbers and observe that $\|x^N\|^2_{\mathcal{C}^N} = \sum_{i=1}^N |\rho_i|^2 \simeq N$.
- Suppose we want to study the algorithm in stationarity and we therefore make the choice $\zeta = 1/3$. With the above point in mind, notice that if we start in stationarity then by the Law of Large numbers $N^{-1} \sum_{i=1}^N |\rho_i|^2 = S^{k,N} \to 1$ (as $N \to \infty$, with speed of convergence $N^{-1/2}$). Moreover, if $x^N \sim \pi_0^N$, by the Central Limit Theorem the term $\langle x^N, \mathcal{C}_N^{1/2} \xi^N \rangle_{\mathcal{C}_N} / \sqrt{N}$ is $O(1)$ and converges to a standard Gaussian. With these two observations in place we can then heuristically see that, with the choice $\zeta = 1/3$ the term in (5.10) are negligible as $N \to \infty$ while the terms in (5.9) are $O(1)$. The term in (5.8) can be better understood by looking at the LHS of (5.11) which, with $\zeta = 1/3$ and $x^N \sim \pi_0^N$, can be rewritten as
$$\frac{\ell^2}{2N^{2/3}} \sum_{i=1}^N (|\rho_i|^2 - |\xi_i|^2). \tag{5.12}$$
The expected value of the above expression is zero. If we apply the Central Limit Theorem to the i.i.d. sequence $\{|\rho_i|^2 - |\xi_i|^2\}_i$, (5.12) shows that (5.8) is $O(N^{1/2-2/3})$ and therefore negligible as $N \to \infty$. In conclusion, in the stationary case the only $O(1)$ terms are those in (5.9); therefore one has the heuristic approximation
$$Q^N(x, \xi) \sim \mathcal{N}\left(-\frac{\ell^3}{4}, \frac{\ell^3}{2} \right).$$
For more details on the stationary case see [PST12].

- If instead we start out of stationarity the choice $\zeta = 1/3$ is problematic. Indeed in [CRR05, Lemma 3] the authors study the MALA algorithm to sample from an N-dimensional isotropic Gaussian and show that if the algorithm is started at a point x^0 such that $S(0) < 1$, then the acceptance probability degenerates to zero. Therefore, the algorithm stays stuck in its initial state and never proceeds to the next move, see [CRR05, Figure 2] (to be more precise, as N increases the algorithm will take longer and longer to get unstuck from its initial state; in the limit, it will never move with probability 1). Therefore the choice $\zeta = 1/3$ cannot be the optimal one (at least not irrespective of the initial state of the chain) if we start out of stationarity. This is still the case in our context and one can heuristically see that the root of the problem lies in the term (5.8). Indeed if out of stationarity we still choose $\zeta = 1/3$ then, like before, (5.9) is still order one and (5.10) is still negligible. However, looking at (5.8), if x^0 is such that $S(0) < 1$ then, when $k = 0$, (5.8) tends to minus infinity; recalling (5.2), this implies that the acceptance probability of the first move tends to zero. To overcome this issue and make Q^N of order one (irrespective of the initial datum) so that the acceptance probability is of order one and does not degenerate to 0 or 1 when $N \to \infty$, we take $\zeta = 1/2$; in this way the terms in (5.8) are $O(1)$, all the others are small. Therefore, the intuition leading the analysis of the non-stationary regime hinges on the fact that, with our scaling,

$$Q^N(x^{k,N}, \xi^{k,N}) \simeq \frac{\ell^2}{2}(S^{k,N} - 1); \tag{5.13}$$

hence

$$\alpha^N(x^{k,N}, \xi^{k,N}) = (1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})}) \simeq \alpha_\ell(S^{k,N}), \tag{5.14}$$

where the function α_ℓ on the RHS of (5.14) is the one defined in (1.11). The approximation (5.13) is made rigorous in Lemma 7.5, while (5.14) is formalized in Section 7.1 (see in particular Proposition 7.4).

- Finally, we mention for completeness that, by arguing similarly to what we have done so far, if $\zeta < 1/2$ then the acceptance probability of the first move tends to zero when $S(0) < 1$. If $\zeta > 1/2$ then $Q^N \to 0$, so the acceptance probability tends to one; however the size of the moves is too small and the algorithm moves in phace space too slowly anyway.

Remark 5.4. Notice that in stationarity the function Q^N is, to leading order, independent of ξ; that is, Q^N and ξ are asymptotically independent (see [PST12, Lemma 4.5]). This can be intuitively explained because in stationarity the leading order term in the expression for Q^N is the term with $\delta^3\|x\|^2$. We will show that also out of stationarity Q^N and ξ are asymptotically independent. In this case such an asymptotic independence can, roughly speaking, be motivated by the approximation (5.13), (as the interpolation of the chain $S^{k,N}$ converges to a deterministic limit). The asymptotic correlation of Q^N and the noise ξ is analysed in Lemma 7.6.

5.2. Heuristic derivation of the weak limit of $S^{k,N}$

Let Y be any function of the random variables $\xi^{k,N}$ and $U^{k,N}$ (introduced in Section 3), for example the chain $x^{k,N}$ itself. Here and throughout the paper we use $\mathbb{E}_{x^0}[Y]$ to denote the expected value of Y with respect to the law of the variables $\xi^{k,N}$'s and $U^{k,N}$'s, with the initial state x_0 of the chain given deterministically; in other words, $\mathbb{E}_{x^0}(Y)$ denotes expectation with respect to all the sources of randomness present in Y. We will use the notation $\mathbb{E}_k[Y]$ for the conditional expectation of Y given $x^{k,N}$, $\mathbb{E}_k[Y] := \mathbb{E}_{x^0}\left[Y \,\middle|\, x^{k,N}\right]$ (we should really be writing $\mathbb{E}_k^N$ in place of $\mathbb{E}_k$, but to improve readability we will omit the further index N). Let us now decompose the chain $S^{k,N}$ into its drift and martingale part:

$$S^{k+1,N} = S^{k,N} + \frac{1}{\sqrt{N}} b_\ell^{k,N} + \frac{1}{N^{1/4}} M^{k,N}, \tag{5.15}$$

where

$$b_\ell^{k,N} := \sqrt{N}\mathbb{E}_k[S^{k+1,N} - S^{k,N}] \tag{5.16}$$

and

$$M^{k,N} := N^{1/4}\left[S^{k+1,N} - S^{k,N} - \frac{1}{\sqrt{N}} b_\ell^{k,N}(x^{k,N})\right] . \tag{5.17}$$

In this subsection we give the heuristics which underly the proof, given in subsequent sections, that the approximate drift $b_\ell^{k,N} = b_\ell^{k,N}(x^{k,N})$ converges to $b_\ell(S^{k,N})$, [3] where b_ℓ is the drift of (1.10), while the approximate diffusion $M^{k,N}$ tends to zero. This formally gives the result of Theorem 5.1. Let us formally argue such a convergence result. By (5.6) and (3.6),

$$S^{k+1,N} = \frac{1}{N}\sum_{j=1}^{N} \frac{\left|x_j^{k+1,N}\right|^2}{\lambda_j^2} = \frac{1}{N}\left(\gamma^{k,N}\left|\left|y^{k,N}\right|\right|_{\mathcal{C}_N}^2 + (1-\gamma^{k,N})\left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^2\right) . \tag{5.18}$$

Therefore, again by (5.6),

$$\begin{aligned} b_\ell^{k,N} = \sqrt{N}\mathbb{E}_k[S^{k+1,N} - S^{k,N}] &= \frac{1}{\sqrt{N}}\mathbb{E}_k\left[\gamma^{k,N}(\left|\left|y^{k,N}\right|\right|_{\mathcal{C}_N}^2 - \left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^2)\right] \\ &= \frac{1}{\sqrt{N}}\mathbb{E}_k\left[(1\wedge e^{Q^N(x^{k,N},y^{k,N})})(\left|\left|y^{k,N}\right|\right|_{\mathcal{C}_N}^2 - \left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^2)\right] , \end{aligned} \tag{5.19}$$

where the second equality is a consequence of the definition of $\gamma^{k,N}$ (with a reasoning, completely analogous to the one in [KOS16, last proof of Appendix A], see also (5.24)). Using (5.3) (with $\delta = \ell/\sqrt{N}$), the fact that r^N is negligible and the approximation (5.13), the above gives

$$b_\ell^{k,N} = \sqrt{N}\mathbb{E}_k[S^{k+1,N} - S^{k,N}] \simeq -\frac{4}{\ell}\left(1\wedge e^{\ell^2(S^{k,N}-1)/2}\right)\frac{\ell^2}{2}(S^{k,N}-1) = b_\ell(S^{k,N}) .$$

The above approximation is made rigorous in Lemma 8.5. As for the diffusion coefficient, it is easy to check (see proof of Lemma 8.2) that

$$N\mathbb{E}_k[S^{k+1,N} - S^{k,N}]^2 < \infty.$$

Hence the approximate diffusion tends to zero and one can formally deduce that (the interpolant of) $S^{k,N}$ converges to the fluid limit (1.10).

5.3. Heuristic analysis of the limit of the chain $x^{k,N}$.

The drift-martingale decomposition of the chain $x^{k,N}$ is as follows:

$$x^{k+1,N} = x^{k,N} + \frac{1}{N^{1/2}}\Theta^{k,N} + \frac{1}{N^{1/4}}L^{k,N} \tag{5.20}$$

where $\Theta^{k,N} = \Theta^{k,N}(x^{k,N})$ is the approximate drift

$$\Theta^{k,N} := \sqrt{N}\mathbb{E}_k\left[x^{k+1,N} - x^{k,N}\right] \tag{5.21}$$

and

$$L^{k,N} := N^{1/4}\left[x^{k+1,N} - x^{k,N} - \frac{1}{\sqrt{N}}\Theta^{k,N}(x^{k,N})\right] \tag{5.22}$$

is the approximate diffusion. In what follows we will use the notation $\Theta(x,S)$ for the drift of equation (1.9), i.e.

$$\Theta(x,S) = F(x)h_\ell(S), \quad (x,S)\in\mathcal{H}^s\times\mathbb{R}, \tag{5.23}$$

with $F(x)$ defined in Lemma 2.5. Again, we want to formally argue that the approximate drift $\Theta^{k,N}(x^{k,N})$ tends to $\Theta(x^{k,N},S^{k,N})$ [4]and the approximate diffusion $L^{k,N}$ tends to the diffusion coefficient of equation (1.9).

[3]Notice that $S^{k,N}$ is only a function of $x^{k,N}$

[4]Note that in the limit the dependence of the drift on $S^{k,N}$ becomes explicit.

5.3.1. Approximate drift.

As a preliminary consideration, observe that

$$\mathbb{E}_k\left(\gamma^{k,N}\mathcal{C}_N^{1/2}\xi^{k,N}\right) = \mathbb{E}_k\left(\left(1\wedge e^{Q^N(x^{k,N},\xi^{k,N})}\right)\mathcal{C}_N^{1/2}\xi^{k,N}\right), \tag{5.24}$$

see [KOS16, equation (5.14)]. This fact will be used throughout the paper, often without mention. Coming to the chain $x^{k,N}$, a direct calculation based on (3.2) and on (3.6) gives

$$x^{k+1,N} - x^{k,N} = -\gamma^{k,N}\delta(x^{k,N} + \mathcal{C}_N\nabla\Psi^N(x^{k,N})) + \gamma^{k,N}\sqrt{2\delta}\mathcal{C}_N^{1/2}\xi^{k,N}. \tag{5.25}$$

Therefore, with the choice $\delta = \ell/\sqrt{N}$, we have

$$\begin{aligned}\Theta^{k,N} = \sqrt{N}\mathbb{E}_k[x^{k+1,N} - x^{k,N}] &= -\ell\mathbb{E}_k\left[(1\wedge e^{Q^N(x^{k,N},\xi^{k,N})})(x^{k,N} + \mathcal{C}_N\nabla\Psi^N(x^{k,N}))\right]\\ &\quad + N^{1/4}\sqrt{2\ell}\mathbb{E}_k\left[(1\wedge e^{Q^N(x^{k,N},\xi^{k,N})})\mathcal{C}_N^{1/2}\,\xi^{k,N}\right]\end{aligned} \tag{5.26}$$

The addend in (5.26) is asymptotically small (see Lemma 7.6 and notice that this addend would just be zero if Q^N and $\xi^{k,N}$ were uncorrelated); hence, using the heuristic approximations (5.13) and (5.14),

$$\begin{aligned}\Theta^{k,N} = \sqrt{N}\mathbb{E}_k[x^{k+1,N} - x^{k,N}] &\simeq -\ell\alpha_\ell(S^{k,N})(x^{k,N} + \mathcal{C}_N\nabla\Psi^N(x^{k,N}))\\ &\overset{(1.12)}{=} -h_\ell(S^{k,N})(x^{k,N} + \mathcal{C}_N\nabla\Psi^N(x^{k,N}));\end{aligned} \tag{5.27}$$

the right hand side of the above is precisely the limiting drift $\Theta(x^{k,N}, S^{k,N})$.

5.3.2. Approximate diffusion.

We now look at the approximate diffusion of the chain $x^{k,N}$:

$$L^{k,N} := N^{1/4}(x^{k+1,N} - x^{k,N} - \mathbb{E}_k(x^{k+1,N} - x^{k,N})).$$

By definition,

$$\mathbb{E}_k\left|\left|L^{k,N}\right|\right|_s^2 = \sqrt{N}\mathbb{E}_k\left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^2 - \sqrt{N}\left|\left|\mathbb{E}_k\left(x^{k+1,N} - x^{k,N}\right)\right|\right|_s^2. \tag{5.28}$$

By (5.27) the second addend in the above is asymptotically small. Therefore

$$\begin{aligned}\mathbb{E}_k\left|\left|L^{k,N}\right|\right|_s^2 &\simeq \sqrt{N}\mathbb{E}_k\left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^2\\ &\overset{(3.6),(5.25)}{\simeq} 2\ell\mathbb{E}_k\left|\left|\gamma^{k,N}\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_s^2\\ &= 2\ell\mathbb{E}_k\sum_{j=1}^N j^{2s}\lambda_j^2\left(1\wedge e^{Q^N(x^{k,N},\xi^{k,N})}\right)\left|\xi_j^{k,N}\right|^2.\end{aligned}$$

The above quantity is carefully studied in Lemma 7.7. However, intuitively, the heuristic approximation (5.14) (and the asymptotic independence of Q^N and ξ that (5.14) is a manifestation of) suffices to formally derive the limiting diffusion coefficient (i.e. the diffusion coefficient of (1.9)):

$$\begin{aligned}\mathbb{E}_k\left|\left|L^{k,N}\right|\right|_s^2 &\simeq 2\ell\sum_{j=1}^N j^{2s}\lambda_j^2\mathbb{E}_k\left[(1\wedge e^{Q^N(x^{k,N},y^{k,N})})\left|\xi_j^{k,N}\right|^2\right]\\ &\simeq 2\ell\sum_{j=1}^N j^{2s}\lambda_j^2\mathbb{E}_k\left[(1\wedge e^{\ell^2(S^{k,N}-1)/2})\left|\xi_j^{k,N}\right|^2\right] \simeq 2\ell\sum_{j=1}^N j^{2s}\lambda_j^2(1\wedge e^{\ell^2(S^{k,N}-1)/2})\\ &\simeq 2\ell\,\mathrm{Trace}(\mathcal{C}_s)\alpha_\ell(S^{k,N}) \overset{(1.12)}{=} 2\mathrm{Trace}(\mathcal{C}_s)\,h_\ell(S^{k,N}).\end{aligned}$$

6. Continuous mapping argument

In this section we outline the argument which underlies the proofs of our main results. In particular, the proofs of Theorem 5.1 and Theorem 5.2 hinge on the continuous mapping arguments that we illustrate in the following Section 6.1 and Section 6.2, respectively. The details of the proofs are deferred to the next three sections: Section 7 contains some preliminary results that we employ in both proofs, Section 8 contains the the proof of Theorem 5.1 and Section 9 that of Theorem 5.2.

6.1. Continuous mapping argument for (4.4)

Let us recall the definition of the chain $\{S^{k,N}\}_{k\in\mathbb{N}}$ and of its continuous interpolant $S^{(N)}$, introduced in (1.14) and (1.15), respectively. From the definition (1.15) of the interpolated process and the drift-martingale decomposition (5.15) of the chain $\{S^{k,N}\}_{k\in\mathbb{N}}$ we have that for any $t \in [t_k, t_{k+1})$,

$$\begin{aligned} S^{(N)}(t) &= (N^{1/2}t - k)\left[S^{k,N} + \frac{1}{\sqrt{N}}b_\ell^{k,N} + \frac{1}{N^{1/4}}M^{k,N}\right] + (k+1-tN^{1/2})S^{k,N} \\ &= S^{k,N} + (t-t_k)b_\ell^{k,N} + N^{1/4}(t-t_k)M^{k,N}. \end{aligned}$$

Iterating the above we obtain

$$S^{(N)}(t) = S^{0,N} + (t-t_k)b_\ell^{k,N} + \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1} b_\ell^{j,N} + w^N(t),$$

where

$$w^N(t) := \frac{1}{N^{1/4}}\sum_{j=0}^{k-1} M^{j,N} + N^{1/4}(t-t_k)M^{k,N} \quad t_k \le t < t_{k+1}. \tag{6.1}$$

The expression for $S^{(N)}(t)$ can then be rewritten as

$$S^{(N)}(t) = S^{0,N} + \int_0^t b_\ell(S^{(N)}(v))dv + \hat{w}^N(t), \tag{6.2}$$

having set

$$\hat{w}^N(t) := e^N(t) + w^N(t), \tag{6.3}$$

with

$$e^N(t) := (t-t_k)b_\ell^{k,N} + \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1} b_\ell^{j,N} - \int_0^t b_\ell(S^{(N)}(v))dv. \tag{6.4}$$

Equation (6.2) shows that

$$S^{(N)} = \mathcal{J}_2(S^{0,N}, \hat{w}^N),$$

where $\mathcal{J}_2$ is the Itô map defined in the statement of Theorem 4.5. By the continuity of the map $\mathcal{J}_2$, if we show that $\hat{w}^N$ converges in probability in $C([0,T];\mathbb{R})$ to zero, then $S^{(N)}(t)$ converges in probability to the solution of the ODE (1.10). We prove convergence of $\hat{w}^N$ to zero in Section 8. In view of (6.3), we show the convergence in probability of $\hat{w}^N$ to zero by proving that both e^N (Lemma 8.1) and w^N (Lemma 8.2) converge in $L_2(\Omega; C([0,T];\mathbb{R}))$ to zero. Because $\{S^{0,N}\}_{N\in\mathbb{N}}$ is a deterministic sequence that converges to S^0, we then have that $(S^{0,N}, \hat{w}^N)$ converges in probability to $(S^0, 0)$.

6.2. Continuous mapping argument for (4.3)

We now consider the chain $\{x^{k,N}\}_{k\in\mathbb{N}} \subseteq \mathcal{H}^s$, defined in (3.8). We act analogously to what we have done for the chain $\{S^{k,N}\}_{k\in\mathbb{N}}$. So we start by recalling the definition of the continuous interpolant $x^{(N)}$, equation (1.8) and the notation introduced at the beginning of Section 5.3. An argument analogous to the one used to derive (6.2) shows that for any $t \in [t_k, t_{k+1})$

$$\begin{aligned} x^{(N)}(t) &= x^{0,N} + (t-t_k)\Theta^{k,N} + \frac{1}{\sqrt{N}}\sum_{j=0}^{k}\Theta^{j,N} + \eta^N(t) \\ &= x^{0,N} + \int_0^t \Theta(x^{(N)}(v), S(v))dv + \hat{\eta}^N(t), \end{aligned} \tag{6.5}$$

where

$$\hat{\eta}^N(t) := d^N(t) + \upsilon^N(t) + \eta^N(t), \tag{6.6}$$

$$\eta^N(t) := N^{1/4}(t-t_k)L^{k,N} + \frac{1}{N^{1/4}}\sum_{j=1}^{k-1} L^{j,N}, \tag{6.7}$$

and

$$d^N(t) := (t-t_k)\Theta^{k,N} + \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1}\Theta^{j,N} - \int_0^t \Theta(x^{(N)}(v), S^{(N)}(v))dv, \tag{6.8}$$

$$\upsilon^N(t) := \int_0^t \left[\Theta(x^{(N)}(v), S^{(N)}(v)) - \Theta(x^{(N)}(v), S(v))\right] dv. \tag{6.9}$$

Equation (6.5) implies that

$$x^{(N)} = \mathcal{J}_1(x^{0,N}, \hat{\eta}^N), \tag{6.10}$$

where $\mathcal{J}_1$ is Itô map defined in the statement of Theorem 4.5. In Section 9 we prove that $\hat{\eta}^N$ converges weakly in $C([0,T];\mathcal{H}^s)$ to the process η, where the process η is the diffusion part of equation (1.9), i.e.

$$\eta(t) := \int_0^t \sqrt{2h_\ell(S(v))}dW_v, \tag{6.11}$$

with W_v a $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion. Looking at (6.6), we prove the weak convergence of $\hat{\eta}^N$ to η by the following steps:

1. We prove that d^N converges in $L_2(\Omega; C([0,T];\mathcal{H}^s))$ to zero (Lemma 9.1);
2. using the convergence in probability (in $C([0,T];\mathbb{R})$) of $S^{(N)}$ to S, we show convergence in probability (in $C([0,T];\mathcal{H}^s)$) of υ^N to zero (Lemma 9.2);
3. we show that η^N converges in weakly in $C([0,T];\mathcal{H}^s)$ to the process η, defined in (6.11) (Lemma 9.3).

Because $\{x^{0,N}\}_{N\in\mathbb{N}}$ is a deterministic sequence that converges to x^0, the above three steps (and Slutsky's Theorem) imply that $(x^{0,N}, \hat{\eta}^N)$ converges weakly to (x^0, η). Now observe that $x(t) = \mathcal{J}_1(x^0, \eta(t))$, where $x(t)$ is the solution of the SDE (4.2). The continuity of the map $\mathcal{J}_1$ (Theorem 4.5), (6.10) and the Continuous Mapping Theorem then imply that the sequence $\{x^{(N)}\}_{N\in\mathbb{N}}$ converges weakly to the solution of the SDE (4.2) (equivalently, to the solution of the SDE (1.9)), thus establishing Theorem 5.2.

7. Preliminary estimates and analysis of the acceptance probability

This section gathers several technical results. In Lemma 7.1 we study the size of the jumps of the chain. Lemma 7.2 contains uniform bounds on the moments of the chains $\{x^{k,N}\}_{k\in\mathbb{N}}$ and $\{S^{k,N}\}_{k\in\mathbb{N}}$, much needed in Section 8 and Section 9. In Section 7.1 we detail the analysis of the acceptance probability. This allows us to quantify the correlations between $\gamma^{k,N}$ and the noise $\xi^{k,N}$, Section 7.2. Throughout the paper, when referring to the function Q^N defined in (5.3), we use interchangeably the notation $Q^N(x^{k,N}, y^{k,N})$ and $Q^N(x^{k,N}, \xi^{k,N})$ (as we have already remarked, given $x^{k,N}$, the proposal $y^{k,N}$ is only a function of $\xi^{k,N}$.)

Lemma 7.1. *Let $q \geq 1/2$ be a real number. Under Assumption 2.1 the following holds:*

$$\mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_s^{2q} \lesssim \frac{1}{N^{q/2}}(1 + \left|\left|x^{k,N}\right|\right|_s^{2q}) \tag{7.1}$$

and

$$\mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_{\mathcal{C}_N}^{2q} \lesssim (S^{k,N})^q + N^{q/2}. \tag{7.2}$$

Therefore,

$$\mathbb{E}_k \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^{2q} \lesssim \frac{1}{N^{q/2}}(1 + \left|\left|x^{k,N}\right|\right|_s^{2q}), \tag{7.3}$$

and

$$\mathbb{E}_k \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_{\mathcal{C}_N}^{2q} \lesssim (S^{k,N})^q + N^{q/2}. \tag{7.4}$$

Proof of Lemma 7.1. By definition of the proposal $y^{k,N}$, equation (3.2),

$$\begin{aligned}\left|\left|y^{k,N} - x^{k,N}\right|\right|_s^{2q} &= \left|\left|\delta(x^{k,N} + \mathcal{C}_N \nabla\Psi^N(x^{k,N})) + \sqrt{2\delta}\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_s^{2q} \\ &\lesssim \frac{1}{N^q}\left(\left|\left|x^{k,N}\right|\right|_s^{2q} + \left|\left|\mathcal{C}_N \nabla\Psi^N(x^{k,N})\right|\right|_s^{2q}\right) + \frac{1}{N^{q/2}}\left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_s^{2q}.\end{aligned}$$

Thus, using (2.15) and (2.17), we have

$$\begin{aligned}\mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_s^{2q} &\lesssim \frac{1}{N^q}\left(1 + \left|\left|x^{k,N}\right|\right|_s^{2q}\right) + \frac{1}{N^{q/2}} \\ &\lesssim \frac{1}{N^{q/2}}\left(1 + \left|\left|x^{k,N}\right|\right|_s^{2q}\right),\end{aligned}$$

which proves (7.1). Equation (7.2) follows similarly:

$$\begin{aligned}\mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_{\mathcal{C}_N}^{2q} &\lesssim \frac{1}{N^q}\left(\left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^{2q} + \left|\left|\mathcal{C}_N \nabla\Psi^N(x^{k,N})\right|\right|_{\mathcal{C}_N}^{2q}\right) \\ &\quad + \frac{1}{N^{q/2}}\mathbb{E}_k \left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_{\mathcal{C}_N}^{2q}.\end{aligned}$$

Since $\left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_{\mathcal{C}_N}^2 = \sum_{j=1}^N (\xi_j^{k,N})^2$ has chi-squared law, applying Stirling's formula for the Gamma function $\Gamma : \mathbb{R} \to \mathbb{R}$ we obtain

$$\mathbb{E}_k \left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_{\mathcal{C}_N}^{2q} \lesssim \frac{\Gamma(q + N/2)}{\Gamma(N/2)} \lesssim N^q. \tag{7.5}$$

Hence, using (2.16), the desired bound follows. Finally, recalling the definition of the chain, equation (3.6), the bounds (7.3) and (7.4) are clearly a consequence of (7.1) and (7.2), respectively, since either $x^{k+1,N} = y^{k,N}$ (if the proposed move is accepted) or $x^{k+1,N} = x^{k,N}$ (if the move is rejected). □

Lemma 7.2. *If Assumption 2.1 holds, then, for every $q \geq 1$, we have*

$$\mathbb{E}_{x^0}(S^{k,N})^q \lesssim 1 \tag{7.6}$$

$$\mathbb{E}_{x^0}\left|\left|x^{k,N}\right|\right|_s^q \lesssim 1, \tag{7.7}$$

uniformly over $N \in \mathbb{N}$ and $k \in \{0, 1, \ldots, [T\sqrt{N}]\}$.

Proof of Lemma 7.2. The proof of this lemma can be found in Appendix B. □

7.1. Acceptance probability

The main result of this section is Proposition 7.4, which we obtain as a consequence of Lemma 7.3 (below) and Lemma 7.2. Proposition 7.4 formalizes the heuristic approximation (5.14).

Lemma 7.3 (Acceptance probability). *Let Assumption 2.1 hold and recall the definitions* (5.2) *and* (1.11). *Then the following holds:*

$$\mathbb{E}_k \left|\alpha^N(x^{k,N},\xi^{k,N}) - \alpha_\ell(S^{k,N})\right|^2 \lesssim \frac{1+(S^{k,N})^2 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}.$$

Before proving Lemma 7.3 , we state Proposition 7.4.

Proposition 7.4. *If Assumption 2.1 holds then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0}\left|\alpha^N(x^{k,N},y^{k,N}) - \alpha_\ell(S^{k,N})\right|^2 = 0.$$

Proof. This is a corollary of Lemma 7.3 and Lemma 7.2. □

Proof of Lemma 7.3. The function $z \mapsto 1 \wedge e^z$ on $\mathbb{R}$ is globally Lipschitz with Lipschitz constant 1. Therefore, by (1.11) and (5.2),

$$\mathbb{E}_k \left|\alpha^N(x^{k,N},y^{k,N}) - \alpha_\ell(S^{k,N})\right|^2 \le \mathbb{E}_k \left|Q^N(x^{k,N},y^{k,N}) - \frac{\ell^2(S^{k,N}-1)}{2}\right|^2.$$

The result is now a consequence of (7.15) below. □

To analyse the acceptance probability it is convenient to decompose Q^N as follows:

$$Q^N(x^N,y^N) = I_1^N(x^N,y^N) + I_2^N(x^N,y^N) + I_3^N(x^N,y^N) \tag{7.8}$$

where

$$\begin{aligned}
I_1^N(x^N,y^N) &:= -\frac{1}{2}\left[\left|\left|y^N\right|\right|_{\mathcal{C}_N}^2 - \left|\left|x^N\right|\right|_{\mathcal{C}_N}^2\right] - \frac{1}{4\delta}\left[\left|\left|x^N-(1-\delta)y^N\right|\right|_{\mathcal{C}_N}^2 - \left|\left|y^N-(1-\delta)x^N\right|\right|_{\mathcal{C}_N}^2\right] \\
&= -\frac{\delta}{4}(\left|\left|y^N\right|\right|_{\mathcal{C}_N}^2 - \left|\left|x^N\right|\right|_{\mathcal{C}_N}^2), && (7.9)\\
I_2^N(x^N,y^N) &:= -\frac{1}{2}\left[\left\langle x^N-(1-\delta)y^N, \mathcal{C}_N\nabla\Psi^N(y^N)\right\rangle_{\mathcal{C}_N} - \left\langle y^N-(1-\delta)x^N, \mathcal{C}_N\nabla\Psi^N(x^N)\right\rangle_{\mathcal{C}_N}\right] \\
&\quad -(\Psi^N(y^N)-\Psi^N(x^N)), && (7.10)\\
I_3^N(x^N,y^N) &:= -\frac{\delta}{4}\left[\left|\left|\mathcal{C}_N\nabla\Psi^N(y^N)\right|\right|_{\mathcal{C}_N}^2 - \left|\left|\mathcal{C}_N\nabla\Psi^N(x^N)\right|\right|_{\mathcal{C}_N}^2\right]. && (7.11)
\end{aligned}$$

Lemma 7.5. *Let Assumption 2.1 hold. With the notation introduced above, we have:*

$$\mathbb{E}_k\left|I_1^N(x^{k,N},y^{k,N}) - \frac{\ell^2(S^{k,N}-1)}{2}\right|^2 \lesssim \frac{\left|\left|x^{k,N}\right|\right|_s^2}{N^2} + \frac{(S^{k,N})^2}{\sqrt{N}} + \frac{1}{N} \tag{7.12}$$

$$\mathbb{E}_k\left|I_2^N(x^{k,N},y^{k,N})\right|^2 \lesssim \frac{1+\left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}} \tag{7.13}$$

$$\mathbb{E}_k\left|I_3^N(x^{k,N},y^{k,N})\right|^2 \lesssim \frac{1}{N}. \tag{7.14}$$

Therefore,

$$\mathbb{E}_k\left|Q^N(x^{k,N},y^{k,N}) - \frac{\ell^2(S^{k,N}-1)}{2}\right|^2 \lesssim \frac{1+(S^{k,N})^2 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}. \tag{7.15}$$

Proof of Lemma 7.5. We consecutively prove the three bounds in the statement.

• **Proof of** (7.12)**.** Using (3.2), we rewrite I_1^N as

$$I_1^N(x^{k,N}, y^{k,N}) = -\frac{\delta}{4}\left(\left|\left|(1-\delta)x^{k,N} - \delta\mathcal{C}_N\nabla\Psi^N(x^{k,N}) + \sqrt{2\delta}\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|^2_{\mathcal{C}_N} - \left|\left|x^{k,N}\right|\right|^2_{\mathcal{C}_N}\right).$$

Expanding the above we obtain:

$$\begin{aligned} I_1^N(x^{k,N}, y^{k,N}) - \frac{\ell^2(S^{k,N}-1)}{2} &= -\left(\frac{\delta^2}{2}\left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|^2_{\mathcal{C}_N} - \frac{\ell^2}{2}\right) \\ &\quad + (r_\Psi^N - r^N) + r_\xi^N + r_x^N, \end{aligned} \tag{7.16}$$

where the difference $(r_\Psi^N - r^N)$ is defined in (5.5) and we set

$$r_\xi^N := -\frac{(\delta^{3/2} - \delta^{5/2})}{\sqrt{2}}\left\langle x^{k,N}, \mathcal{C}_N^{1/2}\xi^{k,N}\right\rangle_{\mathcal{C}_N}, \tag{7.17}$$

$$r_x^N := -\frac{\delta^3}{4}\left|\left|x^{k,N}\right|\right|^2_{\mathcal{C}_N}. \tag{7.18}$$

For the reader's convenience we rearrange (5.5) below:

$$\begin{aligned} r_\Psi^N - r^N = \frac{\delta^2 - \delta^3}{2}&\left\langle x^{k,N}, \mathcal{C}_N\nabla\Psi^N(x^{k,N})\right\rangle_{\mathcal{C}_N} \\ &- \frac{\delta^3}{4}\left|\left|\mathcal{C}_N\nabla\Psi^N(x^{k,N})\right|\right|^2_{\mathcal{C}_N} + \frac{\delta^{5/2}}{\sqrt{2}}\left\langle \mathcal{C}_N\nabla\Psi^N(x^{k,N}), \mathcal{C}_N^{1/2}\xi^{k,N}\right\rangle_{\mathcal{C}_N}. \end{aligned} \tag{7.19}$$

We come to bound all of the above terms, starting from (7.19). To this end, let us observe the following:

$$\left|\left\langle x^{k,N}, \mathcal{C}_N\nabla\Psi^N(x^{k,N})\right\rangle_{\mathcal{C}_N}\right|^2 = \left|\sum_{i=1}^N x_i^{k,N}[\nabla\Psi^N(x^{k,N})]_i\right|^2 \tag{7.20}$$

$$\stackrel{(2.6)}{\leq} \left|\left|x^{k,N}\right|\right|_s^2 \|\nabla\Psi^N(x^{k,N})\|^2_{-s} \stackrel{(2.14)}{\lesssim} \left|\left|x^{k,N}\right|\right|_s^2. \tag{7.21}$$

Moreover,

$$\mathbb{E}_k\left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|^2_{\mathcal{C}_N} = \mathbb{E}_k\sum_{j=1}^N|\xi_j|^2 = N,$$

hence

$$\left|\left\langle \mathcal{C}_N\nabla\Psi^N(x^{k,N}), \mathcal{C}_N^{1/2}\xi^{k,N}\right\rangle_{\mathcal{C}_N}\right|^2 \leq \left|\left|\mathcal{C}_N\nabla\Psi^N(x^{k,N})\right|\right|^2_{\mathcal{C}_N}\left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|^2_{\mathcal{C}_N} \stackrel{(2.16)}{\lesssim} N.$$

From (7.19), (7.20), (2.16) and the above,

$$\mathbb{E}_k\left|r_\Psi^N - r^N\right|^2 \lesssim \frac{\left|\left|x^{k,N}\right|\right|_s^2}{N^2} + \frac{1}{N^{3/2}}. \tag{7.22}$$

By (7.17),

$$\begin{aligned} \mathbb{E}_k\left|r_\xi^N\right|^2 &\lesssim \frac{1}{N^{3/2}}\mathbb{E}_k\left|\left\langle x^{k,N}, \mathcal{C}_N^{1/2}\xi^{k,N}\right\rangle_{\mathcal{C}_N}\right|^2 \\ &= \frac{1}{N^{3/2}}\mathbb{E}_k\left(\sum_{i=1}^N\frac{x_i^{k,N}\xi_i^{k,N}}{\lambda_i}\right)^2 = \frac{1}{\sqrt{N}}S^{k,N}, \end{aligned} \tag{7.23}$$

where in the last equality we have used the fact that $\{\xi_i^{k,N} : i = 1, \dots, N\}$ are independent, zero mean, unit variance normal random variables (independent of $x^{k,N}$) and (5.6). As for r_x^N,

$$\mathbb{E}_k \left|r_x^N\right|^2 \lesssim \frac{1}{N^3} \left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^4 \overset{(5.6)}{=} \frac{(S^{k,N})^2}{N}.$$

Lastly,

$$\tilde{r}^N := \frac{\delta^2}{2} \left|\left|\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_{\mathcal{C}_N}^2 - \frac{\ell^2}{2} = \frac{\ell^2}{2}\left(\frac{1}{N}\sum_{j=1}^N \xi_j^2 - 1\right).$$

Since $\sum_{j=1}^N \xi_j^2$ has chi-squared law, $\mathbb{E}_k \left|\tilde{r}^N\right|^2 \lesssim Var\left(N^{-1}\sum_{j=1}^N \xi_j^2\right) \lesssim N^{-1}$, by (7.5). Combining all of the above, we obtain the desired bound.

- **Proof of** (7.13) From (7.10),

$$\begin{aligned} I_2^N(x^{k,N}, y^{k,N}) = &- \left[\Psi^N(y^{k,N}) - \Psi^N(x^{k,N}) - \left\langle y^{k,N} - x^{k,N}, \nabla\Psi^N(x^{k,N})\right\rangle\right] \\ &+ \frac{1}{2}\left\langle y^{k,N} - x^{k,N}, \nabla\Psi^N(y^{k,N}) - \nabla\Psi^N(x^{k,N})\right\rangle \\ &+ \frac{\delta}{2}\left(\left\langle x^{k,N}, \nabla\Psi^N(x^{k,N})\right\rangle - \left\langle y^{k,N}, \nabla\Psi^N(y^{k,N})\right\rangle\right) =: \sum_{j=1}^3 d_j, \end{aligned}$$

where d_j is the addend on line j of the above array. Using (2.12), (2.14), (2.6) and Lemma 7.1, we have

$$\mathbb{E}_k \left|d_1\right|^2 \lesssim \mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_s^2 \lesssim \frac{1 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}.$$

By the first inequality in (2.14),

$$\left|\left|\nabla\Psi^N(y^{k,N}) - \nabla\Psi^N(x^{k,N})\right|\right|_{-s} \lesssim 1.$$

Consequently, again by (2.6) and Lemma 7.1,

$$\mathbb{E}_k \left|d_2\right|^2 \lesssim \mathbb{E}_k \left|\left|y^{k,N} - x^{k,N}\right|\right|_s^2 \lesssim \frac{1 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}.$$

Next, applying (2.6) and (2.14) gives

$$\begin{aligned} |d_3| &\leq \frac{\left|\left|x^{k,N}\right|\right|_s \left|\left|\nabla\Psi^N(x^{k,N})\right|\right|_{-s} + \left|\left|y^{k,N}\right|\right|_s \left|\left|\nabla\Psi^N(y^{k,N})\right|\right|_{-s}}{\sqrt{N}} \\ &\lesssim \frac{\left|\left|x^{k,N}\right|\right|_s + \left|\left|y^{k,N}\right|\right|_s}{\sqrt{N}} \lesssim \frac{\left|\left|x^{k,N}\right|\right|_s + \left|\left|y^{k,N} - x^{k,N}\right|\right|_s}{\sqrt{N}}. \end{aligned}$$

Thus, applying Lemma 7.1 then gives the desired bound.

- **Proof of** (7.14) This follows directly from (2.15).

□

7.2. Correlations between the acceptance probability and the noise $\xi^{k,N}$

Recall the definition of $\gamma^{k,N}$, equation (3.7), and let

$$\varepsilon^{k,N} := \gamma^{k,N}\mathcal{C}_N^{1/2}\xi^{k,N}. \tag{7.24}$$

The study of the properties of $\varepsilon^{k.N}$ is the object of the next two lemmata, which have a central role in the analysis: Lemma 7.6 (and Lemma 7.2) establishes the decay of correlations between the acceptance probability and the noise $\xi^{k,N}$. Lemma 7.7 formalizes the heuristic arguments presented in Section 5.3.2.

Lemma 7.6. *If Assumption 2.1 holds, then*

$$\left|\left|\mathbb{E}_k \varepsilon^{k,N}\right|\right|_s^2 \lesssim \frac{1+\left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}. \tag{7.25}$$

Therefore,

$$\left\langle \mathbb{E}_k \varepsilon^{k,N}, x^{k,N}\right\rangle_s = \mathbb{E}_k \left\langle \gamma^{k,N} \mathcal{C}_N^{1/2} \xi^{k,N}, x^{k,N}\right\rangle_s \lesssim \frac{1}{N^{1/4}}(1+\left|\left|x^{k,N}\right|\right|_s^2). \tag{7.26}$$

Lemma 7.7. *Let Assumption 2.1 hold. Then, with the notation introduced so far,*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left|\mathbb{E}_k \left|\left|\varepsilon^{k,N}\right|\right|_s^2 - \mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\alpha_\ell(S^{k,N})\right| = 0.$$

The proofs of the above lemmata can be found in Appendix A. Notice that if $\xi^{k,N}$ and $\gamma^{k,N}$ (equivalently $\xi^{k,N}$ and Q^N) were uncorrelated, the statements of Lemma 7.6 and Lemma 7.7 would be trivially true.

8. Proof of Theorem 5.1

As explained in Section 6.1, due to the continuity of the map $\mathcal{J}_2$ (defined in Theorem 4.5), in order to prove Theorem 5.1 all we need to show is convergence in probability of $\hat{w}^N(t)$ to zero. Looking at the definition of $\hat{w}^N(t)$, equation (6.3), the convergence in probability (in $C([0,T];\mathbb{R})$) of $\hat{w}^N(t)$ to zero is consequence of Lemma 8.1 and Lemma 8.2 below. We prove Lemma 8.1 in Section 8.1 and Lemma 8.2 in Section 8.2.

Lemma 8.1. *Let Assumption 2.1 hold and recall the definition (6.4) of the process $e^N(t)$; then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|e^N(t)\right|\right)^2 = 0.$$

Lemma 8.2. *Let Assumption 2.1 hold and recall the definition (6.1) of the process $w^N(t)$; then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|w^N(t)\right|\right)^2 = 0.$$

8.1. Analysis of the drift

In view of what follows, it is convenient to introduce the piecewise constant interpolant of the chain $\{S^{k,N}\}_{k\in\mathbb{N}}$:

$$\bar{S}^{(N)}(t) := S^{k,N}, \quad t_k \le t < t_{k+1}, \tag{8.1}$$

where $t_k = k/\sqrt{N}$.

Proof of Lemma 8.1. From (8.1), for any $t_k \le t < t_{k+1}$ we have

$$\begin{aligned}\int_0^t b_\ell(\bar{S}_v^{(N)})dv &= \int_{t_k}^t b_\ell(\bar{S}_v^{(N)})dv + \sum_{j=1}^{k-1}\int_{t_{j-1}}^{t_j} b_\ell(\bar{S}_v^{(N)})dv \\ &= (t-t_k)b_\ell(S^{k,N}) + \frac{1}{\sqrt{N}}\sum_{j=1}^{k-1} b_\ell(S^{j,N}).\end{aligned}$$

With this observation, we can then decompose $e^N(t)$ as

$$e^N(t) = e_1^N(t) - e_2^N(t),$$

where

$$e_1^N(t) := (t-t_k)(b_\ell^{k,N} - b_\ell(S^{k,N})) + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \left[b_\ell^{j,N} - b_\ell(S^{j,N})\right] \tag{8.2}$$

$$e_2^N(t) := \int_0^t \left[b_\ell(S_v^{(N)}) - b_\ell(\bar{S}_v^{(N)})\right] dv. \tag{8.3}$$

The result is now a consequence of Lemma 8.3 and Lemma 8.4 below, which we first state and then consecutively prove. □

Lemma 8.3. *If Assumption 2.1 holds, then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|e_1^N(t)\right| \right)^2 = 0.$$

Lemma 8.4. *If Assumption 2.1 holds, then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|e_2^N(t)\right| \right)^2 = 0.$$

Proof of Lemma 8.3. Denoting $E^{k,N} := b_\ell^{k,N} - b_\ell(S^{k,N})$, by (discrete) Jensen's inequality we have

$$\begin{aligned}\sup_{t\in[0,T]} \left|e_1^N(t)\right|^2 &= \sup_{t\in[0,T]} \left|(t-t_k)E^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} E^{k,N}\right|^2 \\ &\lesssim \frac{1}{\sqrt{N}} \sum_{j=0}^{[T\sqrt{N}]-1} \left|E^{j,N}\right|^2.\end{aligned}$$

Using Lemma 8.5 below, we obtain

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{[T\sqrt{N}]-1} \left|E^{j,N}\right|^2 \lesssim \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]-1} \frac{1 + (S^{k,N})^4 + \left|\left|x^{k,N}\right|\right|_s^4}{\sqrt{N}}.$$

Taking expectations on both sides and applying Lemma 7.2 completes the proof. □

Lemma 8.5. *Let Assumption 2.1 hold. Then, for any $N \in \mathbb{N}$ and $k \in \{0, 1, \dots, [T\sqrt{N}]\}$,*

$$\left|E^{k,N}\right|^2 = \left|b_\ell^{k,N} - b_\ell(S^{k,N})\right|^2 \lesssim \frac{1 + (S^{k,N})^4 + \left|\left|x^{k,N}\right|\right|_s^4}{\sqrt{N}}.$$

Proof of Lemma 8.5. Define

$$Y_k^N := \frac{\left|\left|y^{k,N}\right|\right|_{\mathcal{C}_N}^2 - \left|\left|x^{k,N}\right|\right|_{\mathcal{C}_N}^2}{\sqrt{N}}, \qquad \tilde{Y}_k^N := 2\ell(1 - S^{k,N}).$$

Then, from (5.19), (5.2), (1.11) and (1.13), we obtain

$$\begin{aligned}\left|b_\ell^{k,N} - b_\ell(S^{k,N})\right|^2 &= \left|\mathbb{E}_k \left(\alpha^N(x^{k,N}, y^{k,N}) Y_k^N\right) - \alpha_\ell(S^{k,N}) \tilde{Y}_k^N\right|^2 \\ &\leq \mathbb{E}_k \left|\alpha^N(x^{k,N}, y^{k,N}) Y_k^N - \alpha_\ell(S^{k,N}) \tilde{Y}_k^N\right|^2 \\ &\lesssim \mathbb{E}_k \left[\left|\alpha^N(x^{k,N}, y^{k,N})\right|^2 \left|Y_k^N - \tilde{Y}_k^N\right|^2\right] \\ &\quad + \mathbb{E}_k \left[\left|\tilde{Y}_k^N\right|^2 \left|\alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N})\right|^2\right].\end{aligned}$$

Since $|\alpha^N(x^{k,N},y^{k,N})|\leq 1$ and $\tilde{Y}_k^N$ is a function of $x^{k,N}$ only, we can further estimate the above as follows:

$$\left|b_\ell^{k,N}-b_\ell(S^{k,N})\right|^2\lesssim \mathbb{E}_k\left|Y_k^N-\tilde{Y}_k^N\right|^2+\left|\tilde{Y}_k^N\right|^2\mathbb{E}_k\left|\alpha^N(x^{k,N},y^{k,N})-\alpha_\ell(S^{k,N})\right|^2. \tag{8.4}$$

From the definition of I_1^N, equation (7.9), we have

$$Y^{k,N}=-\frac{4}{\ell}I_1^N(x^{k,N},y^{k,N}). \tag{8.5}$$

Therefore,

$$Y_k^N-\tilde{Y}_k^N=-\frac{4}{\ell}\left[I_1^N-\frac{\ell^2}{2}(S^{k,N}-1)\right],$$

which implies

$$\mathbb{E}_k(Y_k^N-\tilde{Y}_k^N)^2\lesssim\mathbb{E}_k\left(I_1^N(x^{k,N},y^{k,N})-\ell^2(S^{k,N}-1)/2\right)^2\overset{(7.12)}{\lesssim}\frac{\left|\left|x^{k,N}\right|\right|_s^2}{N^2}+\frac{(S^{k,N})^2}{\sqrt{N}}+\frac{1}{N}.$$

As for the second addend in (8.4), Lemma 7.3 gives

$$\begin{aligned}\left|\tilde{Y}_k^N\right|^2\mathbb{E}_k\left|\alpha^N(x^{k,N},y^{k,N})-\alpha_\ell(S^{k,N})\right|^2&\lesssim(1+(S^{k,N})^2)\left(\frac{1+(S^{k,N})^2+\left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}}\right)\\&\lesssim\frac{1+(S^{k,N})^4+\left|\left|x^{k,N}\right|\right|_s^4}{\sqrt{N}}.\end{aligned}$$

Combining the above two bounds and (8.4) gives the desired result. □

Proof of Lemma 8.4. By Jensen's inequality,

$$\left(\sup_{t\in[0,T]}\left|\int_0^t b_\ell(S_v^{(N)})-b_\ell(\bar{S}_v^{(N)})dv\right|\right)^2\lesssim\int_0^T\left|b_\ell(S_v^{(N)})-b_\ell(\bar{S}_v^{(N)})\right|^2dv.$$

Since b_ℓ is globally Lipschitz,

$$\begin{aligned}\int_0^T\left|b_\ell(\bar{S}^N(v))-b_\ell(S^N(v))\right|^2dv&\lesssim\int_0^T\left|\bar{S}^N(v)-S^N(v)\right|^2dv\\&=\sum_{k=0}^{[T\sqrt{N}]-1}\int_{t_k}^{t_{k+1}}\left|\bar{S}^N(v)-S^N(v)\right|^2dv+\int_{[T\sqrt{N}]}^T\left|\bar{S}^N(v)-S^N(v)\right|^2dv\\&\lesssim\frac{1}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]-1}(S^{k+1,N}-S^{k,N})^2.\end{aligned}$$

From (5.18) and (5.6),

$$\begin{aligned}\left|S^{k+1,N}-S^{k,N}\right|&\lesssim\frac{1}{N}\left(\|y^{k,N}\|_{\mathcal{C}^N}^2-\|x^{k,N}\|_{\mathcal{C}^N}^2\right)\\&\overset{(8.5)}{\lesssim}\frac{1}{\sqrt{N}}I_1^N(x^{k,N},y^{k,N})\\&=\frac{1}{\sqrt{N}}\left(I_1^N(x^{k,N},y^{k,N})-\frac{\ell^2(S^{k,N}-1)}{2}\right)+\frac{1}{\sqrt{N}}\frac{\ell^2(S^{k,N}-1)}{2}.\end{aligned}$$

Combining the above with (7.12) we obtain

$$\mathbb{E}_k(S^{k+1,N}-S^{k,N})^2\lesssim\frac{1+(S^{k,N})^2+\left|\left|x^{k,N}\right|\right|_s^2}{N}. \tag{8.6}$$

Taking expectations and applying Lemma 7.2 concludes the proof. □

8.2. Analysis of the noise

Proof of Lemma 8.2. After a calculation analogous to the one at the beginning of the proof of Lemma 8.3, all we need to prove is the following limit:

$$\frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_{x^0} \left| M^{k,N} \right|^2 \to 0 \quad \text{as} \quad N \to \infty.$$

By the definition of $M^{k,N}$, equation (5.17), we have

$$\begin{aligned}\frac{\mathbb{E}_{x^0} \left| M^{k,N} \right|^2}{\sqrt{N}} &= \mathbb{E}_{x^0} \left[S^{k+1,N} - S^{k,N} - \mathbb{E}_k \left(S^{k+1,N} - S^{k,N} \right) \right]^2 \\ &\lesssim \mathbb{E}_{x^0} \left| S^{k+1,N} - S^{k,N} \right|^2 \lesssim \frac{1}{N},\end{aligned}$$

where the last inequality is a consequence of (8.6) and Lemma 7.2. This concludes the proof. □

9. Proof of Theorem 5.2

The idea behind the proof is the same as in the previous Section 8. First we introduce the piecewise constant interpolant of the chain $\{x^{k,N}\}_{k\in\mathbb{N}}$

$$\bar{x}^{(N)}(t) = x^{k,N} \quad \text{for} \quad t_k \le t < t_{k+1}. \tag{9.1}$$

Due to the continuity of the map $\mathcal{J}_1$ (Theorem 4.5), all we need to prove is the weak convergence of $\hat{\eta}^N(t)$ to zero (see Section 6.2). Looking at the definition of $\hat{\eta}^N(t)$, equation (6.6), this follows from Lemmas 9.1, 9.2 and 9.3 below. We prove Lemma 9.1 and Lemma 9.2 in Section 9.1 and Lemma 9.3 in Section 9.2.

Lemma 9.1. *Let Assumption 2.1 hold and recall the definition (6.8) of the process $d^N(t)$; then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left| d^N(t) \right| \right)^2 = 0.$$

Lemma 9.2. *If Assumption 2.1 holds, then υ^N (defined in (6.9)) converges in probability in $C([0,T];\mathcal{H}^s)$ to zero.*

Lemma 9.3. *Let Assumption 2.1 hold. Then the interpolated martingale difference array $\eta^N(t)$ defined in (6.7) converges weakly in $C([0,T];\mathcal{H}^s)$ to the stochastic integral $\eta(t)$, defined in equation (6.11).*

9.1. Analysis of the drift

Proof of Lemma 9.1. For all $t \in [t_k, t_{k+1})$, we can write

$$(t - t_k)\Theta(x^{k,N}, S^{k,N}) + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \Theta(x^{j,N}, S^{j,N}) = \int_0^t \Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v)) dv.$$

Therefore, we can decompose $d^N(t)$ as

$$d^N(t) = d_1^N(t) + d_2^N(t),$$

where

$$d_1^N(t) := (t - t_k)\left[\Theta^{k,N} - \Theta(x^{k,N}, S^{k,N})\right] + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \left[\Theta^{j,N} - \Theta(x^{j,N}, S^{j,N})\right]$$

and

$$d_2^N(t) := \int_0^t \left[\Theta(\bar{x}^N(v), \bar{S}^N(v)) - \Theta(x^{(N)}(v), S^{(N)}(v))\right] dv.$$

The statement is now a consequence of Lemma 9.4 and Lemma 9.5. □

Lemma 9.4. *If Assumption 2.1 holds, then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|\left|d_1^N(t)\right|\right|_s \right)^2 = 0.$$

Lemma 9.5. *If Assumption 2.1 holds, then*

$$\lim_{N\to\infty} \mathbb{E}_{x^0} \left(\sup_{t\in[0,T]} \left|\left|d_2^N(t)\right|\right|_s \right)^2 = 0.$$

Before proving Lemma 9.4, we state and prove the following Lemma 9.6. We then consecutively prove Lemma 9.4, Lemma 9.5 and Lemma 9.2. Recall the definitions of Θ and $\Theta^{k,N}$, equations (5.23) and (5.21), respectively.

Lemma 9.6. *Let Assumption 2.1 hold and set*

$$p^{k,N} := \Theta^{k,N} - \Theta(x^{k,N}, S^{k,N}). \tag{9.2}$$

Then

$$\mathbb{E}_{x^0} \left|\left|p^{k,N}\right|\right|_s^2 \lesssim \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4 + \frac{1}{\sqrt{N}}.$$

Proof of Lemma 9.6. Recalling (5.26) and (7.24), we have

$$\left|\left|p^{k,N}\right|\right|_s^2 \lesssim \sqrt{N} \left|\left|\mathbb{E}_k \varepsilon_k^N(x^{k,N})\right|\right|_s^2 \tag{9.3}$$

$$+ \left|\left|\alpha_\ell(S^{k,N}) F(x^{k,N}) - \left[\mathbb{E}_k \alpha^N(x^{k,N}, y^{k,N})\right] (x^{k,N} + \mathcal{C}_N \nabla\Psi^N(x^{k,N}))\right|\right|_s^2, \tag{9.4}$$

where the function F that appears in the above has been defined in Lemma 2.5. The term on the RHS of (9.3) has been studied in Lemma 7.6. To estimate the addend in (9.4) we use (2.15), the boundedness of α_ℓ and Lemma 7.3. A straightforward calculation then gives

$$\begin{aligned}
(9.4) &\lesssim \left[\alpha_\ell(S^{k,N}) - \mathbb{E}_k \alpha^N(x^{k,N}, y^{k,N})\right]^2 \left|\left|(x^{k,N} + \mathcal{C}_N \nabla\Psi^N(x^{k,N}))\right|\right|_s^2 \\
&\quad + \left|\left|\alpha_\ell(S^{k,N}) \left[F(x^{k,N}) - (x^{k,N} + \mathcal{C}_N \nabla\Psi^N(x^{k,N}))\right]\right|\right|_s^2 \\
&\lesssim \frac{1 + (S^{k,N})^4 + \left|\left|x^{k,N}\right|\right|_s^4}{\sqrt{N}} + \left|\left|\mathcal{C}\nabla\Psi(x^{k,N}) - \mathcal{C}_N \nabla\Psi^N(x^{k,N})\right|\right|_s^2.
\end{aligned}$$

From the definition of Ψ^N and $\nabla\Psi^N$, equation (1.5) and equation (2.13), respectively,

$$\begin{aligned}
\left|\left|\mathcal{C}\nabla\Psi(x^{k,N}) - \mathcal{C}_N \nabla\Psi^N(x^{k,N})\right|\right|_s^2 &= \left|\left|\mathcal{C}\nabla\Psi(x^{k,N}) - \mathcal{C}_N \mathcal{P}^N(\nabla\Psi(x^{k,N}))\right|\right|_s^2 \\
&= \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4 \mathbb{E}\left[j^{-2s} (\nabla\Psi(x^{k,N}))_j^2\right] \lesssim \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4,
\end{aligned}$$

having used (2.14) in the last inequality. The statement is now a consequence of Lemma 7.2. □

Proof of Lemma 9.4. Following the analogous steps to those taken in the proof of Lemma 8.3, the proof is a direct consequence of Lemma 9.6, after observing that the summation $\sum_{j=N+1}^{\infty}(\lambda_j j^s)^4$ is the tail of a convergent series hence it tends to zero as $N \to \infty$. $\square$

Proof of Lemma 9.5. By the definition of Θ, equation (5.23), we have

$$\left|\left|\Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t))\right|\right|_s = \left|\left|F(\bar{x}^N)h_\ell(\bar{S}^N) - F(x^{(N)})h_\ell(S^{(N)})\right|\right|_s.$$

Applying (2.10) and (2.15) and using the fact h_ℓ is globally Lipschitz and bounded, we get

$$\left|\left|\Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t))\right|\right|_s \lesssim \left|\left|\bar{x}^N(t) - x^{(N)}(t)\right|\right|_s + (1 + \left|\left|\bar{x}^N(t)\right|\right|_s)\left|\bar{S}^N(t) - S^{(N)}(t)\right|.$$

Thus, from the definitions (1.15), (8.1), (1.8) and (9.1), if $t_k \le t < t_{k+1}$, we have

$$\begin{aligned}\left|\left|\Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t))\right|\right|_s &\lesssim (t - k\sqrt{N})\left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s \\ &\quad + (t - k\sqrt{N})(1 + \left|\left|x^{k,N}\right|\right|_s)\left|S^{k+1,N} - S^{k,N}\right|.\end{aligned}$$

Applying (7.3) and (8.6) one then concludes

$$\mathbb{E}_k\left|\left|\Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t))\right|\right|_s^2 \lesssim (t - k\sqrt{N})^2\left(\frac{1 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}} + \frac{\left|\left|x^{k,N}\right|\right|_s^4 + (S^{k,N})^4}{N}\right)$$

The remainder of the proof is analogous to the proof of Lemma 8.4. $\square$

Proof of Lemma 9.2. For any arbitrary but fixed $\varepsilon > 0$, we need to argue that

$$\lim_{N\to\infty}\mathbb{P}\left[\sup_{t\in[0,T]}\left|\left|\upsilon^N(t)\right|\right|_s \ge \varepsilon\right] = 0.$$

From the definition of υ^N we have

$$\sup_{t\in[0,T]}\left|\left|\upsilon^N(t)\right|\right|_s \le \int_0^T\left|\left|F(x^{(N)}(v))\right|\right|_s\left|S^{(N)}(v) - S(v)\right|dv.$$

Using (2.11) and the fact that $\left|\left|x^{(N)}(t)\right|\right|_s \le \left|\left|x^{k,N}\right|\right|_s + \left|\left|x^{k+1,N}\right|\right|_s$ (which is a simple consequence of (1.8)), for any $t \in [t_k, t_{k+1})$

$$\begin{aligned}\sup_{t\in[0,T]}\left|\left|\upsilon^N(t)\right|\right|_s &\le \left(\sup_{t\in[0,T]}\left|S^{(N)}(t) - S(t)\right|\right)\int_0^T\left|\left|F(x^{(N)}(v))\right|\right|_s dv \\ &\lesssim \underbrace{\left(\sup_{t\in[0,T]}\left|S^{(N)}(t) - S(t)\right|\right)}_{=:a^N}\underbrace{\left(1 + \frac{1}{\sqrt{N}}\sum_{j=0}^{[T\sqrt{N}]-1}\left|\left|x^{j,N}\right|\right|_s\right)}_{=:u^N}.\end{aligned}$$

Using Markov's inequality and Lemma 7.2, given any $\delta > 0$, it is straightforward to find constant M such that $\mathbb{P}\left[u^N > M\right] \le \delta$ for every $N \in \mathbb{N}$. Thus

$$\begin{aligned}\mathbb{P}\left[\sup_{t\in[0,T]}\left|\left|\upsilon^N(t)\right|\right|_s \ge \varepsilon\right] &\le \mathbb{P}\left[a^N u^N \ge \varepsilon\right] = \mathbb{P}[a^N u^N \ge \varepsilon, u^N \le M] + \mathbb{P}[a^N u^N \ge \varepsilon, u^N > M] \\ &\le \mathbb{P}\left[a^N \ge \varepsilon/M\right] + \mathbb{P}\left[u^N > M\right] \le \mathbb{P}\left[a^N \ge \varepsilon/M\right] + \delta.\end{aligned}$$

Given that the δ was arbitrary, the result then follows from the fact that $S^{(N)}$ converges in probability to S (Theorem 5.1). $\square$

9.2. Analysis of the noise

The proof of Lemma 9.3 is based on [KOS16, Lemma 8.9]. For the reader's convenience, we restate [KOS16, Lemma 8.9] below as Lemma 9.7. In order to state such a lemma let us introduce the following notation and definitions. Let $k_N : [0,T] \to \mathbb{Z}_+$ be a sequence of nondecreasing, right continuous functions indexed by N, with $k_N(0) = 0$ and $k_N(T) \geq 1$. Let $\mathcal{H}$ be any Hilbert space and $\{X^{k,N}, \mathcal{F}^{k,N}\}_{0\leq k\leq k_N(T)}$ be a $\mathcal{H}$-valued martingale difference array (MDA), i.e. a double sequence of random variables such that $\mathbb{E}[X^{k,N}|\mathcal{F}^N_{k-1}] = 0$, $\mathbb{E}[\|X^{k,N}\|^2|\mathcal{F}^N_{k-1}] < \infty$ almost surely and sigma-algebras $\mathcal{F}^{k-1,N} \subseteq \mathcal{F}^{k,N}$. Consider the process $\mathcal{X}^N(t)$ defined by

$$\mathcal{X}^N(t) := \sum_{k=1}^{k_N(t)} X^{k,N},$$

if $k_N(t) \geq 1$ and $k_N(t) > \lim_{v\to 0+} k_N(t-v)$ and by linear interpolation otherwise. With this set up we recall the following result.

Lemma 9.7 (Lemma 8.9 [KOS16])**.** *Let $D : \mathcal{H} \to \mathcal{H}$ be a self-adjoint positive definite trace class operator on $(\mathcal{H}, ||\cdot||)$. Suppose the following limits hold in probability*

i) *there exists a continuous and positive function $f : [0,T] \to \mathbb{R}_+$ such that*

$$\lim_{N\to\infty} \sum_{k=1}^{k_N(T)} \mathbb{E}(\left|\left|X^{k,N}\right|\right|^2 |\mathcal{F}^N_{k-1}) = \mathrm{Trace}_{\mathcal{H}}(D) \int_0^T f(t)dt\,;$$

ii) *if $\{\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$ then*

$$\lim_{N\to\infty} \sum_{k=1}^{k_N(T)} \mathbb{E}(\langle X^{k,N}, \phi_j\rangle\langle X^{k,N}, \phi_i\rangle |\mathcal{F}^N_{k-1}) = 0 \quad \textit{for all} \;\; i \neq j\,;$$

iii) *for every fixed $\epsilon > 0$,*

$$\lim_{N\to\infty} \sum_{k=1}^{k_N(T)} \mathbb{E}(\left|\left|X^{k,N}\right|\right|^2 \mathbf{1}_{\left\{||X^{k,N}||^2\geq\epsilon\right\}} |\mathcal{F}^N_{k-1}) = 0, \qquad \textit{in probability},$$

where $\mathbf{1}_A$ denotes the indicator function of the set A. Then the sequence $\mathcal{X}^N$ converges weakly in $C([0,T];\mathcal{H}^s)$ to the stochastic integral $t \mapsto \int_0^t \sqrt{f(v)}dW_v$, where W_t is a $\mathcal{H}$-valued D-Brownian motion.

Proof of Lemma 9.3. We apply Lemma 9.7 in the Hilbert space $\mathcal{H}^s$, with $k_N(t) = [t\sqrt{N}]$, $X^{k,N} = L^{k,N}/N^{1/4}$ ($L^{k,N}$ is defined in (5.22)) and $\mathcal{F}^N_k$ the sigma-algebra generated by $\{\gamma^{h,N}, \xi^{h,N}, 0 \leq h \leq k\}$ to study the sequence $\eta^N(t)$, defined in (6.7). We now check that the three conditions of Lemma 9.7 hold in the present case.

i) Note that by the definition of $L^{k,N}$, $\mathbb{E}[L^{k,N}|\mathcal{F}^N_{k-1}] = \mathbb{E}_k[L^{k,N}]$ almost surely.We need to show that the limit

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left|\left|L^{k,N}\right|\right|_s^2 = 2\,\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) \int_0^T h_\ell(S(u))du\,, \tag{9.5}$$

holds in probability. By (5.28),

$$\frac{1}{\sqrt{N}} \mathbb{E}_k \left|\left|L^{k,N}\right|\right|_s^2 = \mathbb{E}_k \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^2 - \left|\left|\mathbb{E}_k \left(x^{k+1,N} - x^{k,N}\right)\right|\right|_s^2.$$

From the above, if we prove

$$\mathbb{E}_{x^0} \sum_{k=0}^{[T\sqrt{N}]} \left|\left|\mathbb{E}_k \left(x^{k+1,N} - x^{k,N}\right)\right|\right|_s^2 \to 0 \quad \text{as } N \to \infty, \tag{9.6}$$

and that

$$\lim_{N\to\infty}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|x^{k+1,N}-x^{k,N}\right|\right|_s^2 = 2\,\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\int_0^T h_\ell(S(u))du, \quad \text{in probability,} \tag{9.7}$$

then (9.5) follows. We start by proving (9.6):

$$\begin{aligned}\left|\left|\mathbb{E}_k\left(x^{k+1,N}-x^{k,N}\right)\right|\right|_s^2 &\overset{(3.8)}{\lesssim} \left|\left|x^{k,N}+\mathcal{C}_N\nabla\Psi^N(x^{k,N})\right|\right|_s^2+\frac{1}{\sqrt{N}}\left|\left|\mathbb{E}_k\left(\gamma^{k,N}(\mathcal{C}_N)^{1/2}\xi^{k,N}\right)\right|\right|_s^2\\ &\lesssim \frac{1}{N}\left(1+\left|\left|x^{k,N}\right|\right|_s^2\right),\end{aligned}$$

where the last inequality follows from (2.15) and (7.25).The above and (7.7) prove (9.6). We now come to (9.7):

$$\begin{aligned}&\left|\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|x^{k+1,N}-x^{k,N}\right|\right|_s^2-2\,\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\int_0^T h_\ell(S(u))du\right|\\ &\overset{(3.8)}{\lesssim}\frac{1}{N}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|x^{k,N}+\mathcal{C}_N\nabla\Psi^N(x^{k,N})\right|\right|_s^2\\ &+\frac{1}{N^{3/4}}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\langle x^{k,N}+\mathcal{C}_N\nabla\Psi^N(x^{k,N}),\mathcal{C}_N^{1/2}\xi^{k,N}\rangle_s\right|\\ &+\left|\frac{2\ell}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|\gamma^{k,N}\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_s^2-2\,\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\int_0^T h_\ell(S(u))du\right|.\end{aligned}$$

The first two addends tend to zero in L_1 as N tends to infinity due to (2.15), (2.17) and Lemma 7.2. As for the third addend, we decompose it as follows

$$\begin{aligned}&\left|\frac{2\ell}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|\gamma^{k,N}\mathcal{C}_N^{1/2}\xi^{k,N}\right|\right|_s^2-2\,\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\int_0^T h_\ell(S(u))du\right|\\ &\overset{(1.12),(7.24)}{\lesssim}\left|\frac{\ell}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}\mathbb{E}_k\left|\left|\varepsilon^{k,N}\right|\right|_s^2-\frac{\ell}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\alpha_\ell(S^{k,N})\right|\\ &+\left|\frac{1}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)h_\ell(S^{k,N})-\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\int_0^T h_\ell(S(u))du\right|.\end{aligned} \tag{9.8}$$

The first addend in the above tends to zero in L_1 due to Lemma 7.7. As for the term in (9.8), we use the identity

$$\int_0^T h_\ell(\bar{S}^{(N)}(u))du=\left(T-\frac{[T\sqrt{N}]}{\sqrt{N}}\right)h_\ell(S^{[T\sqrt{N}],N})+\frac{1}{\sqrt{N}}\sum_{k=0}^{[T\sqrt{N}]}h_\ell(S^{k,N}),$$

to further split it, obtaining:

$$(9.8) \lesssim \left| \int_0^T h_\ell(\bar{S}^{(N)}(u)) - h_\ell(S^{(N)}(u)) du \right| \tag{9.9}$$

$$+ \left| \int_0^T h_\ell(S^{(N)}(u)) - h_\ell(S(u)) du \right| \tag{9.10}$$

$$+ \left(T - \frac{[T\sqrt{N}]}{\sqrt{N}} \right) h_\ell(S^{[T\sqrt{N}],N}). \tag{9.11}$$

Convergence (in L_1) of (9.9) to zero follows with the same calculations leading to (8.6), the global Lipschitz property of h_ℓ, and Lemma 7.2. The addend in (9.10) tends to zero in probability since $S^{(N)}$ tends to S in probability in $C([0,T];\mathbb{R})$ (Theorem 5.1) and the third addend is clearly small. The limit (9.7) then follows.

ii) Condition **ii)** of Lemma 9.7 can be shown to hold with similar calculations, so we will not show the details.

iii) Using (7.3) , the last bound follows a calculation completely analogous to the one in [KOS16, Section 8.2] so we don't repeat details here.

□

Appendix A: Proofs of results of Section 7.2

In view of the proof of Lemma 7.6 and Lemma 7.7, let us decompose $Q^N(x^{k,N}, \xi^{k,N})$ into a term that depends on $\xi_j^{k,N}$ (the j-th component of $\xi^{k,N}$), Q_j^N, and a term that is independent of ξ_j, $Q_{j,\perp}^N$:

$$Q^N(x, \xi) = Q_j^N + Q_{j,\perp}^N,$$

where

$$\begin{aligned} Q_j^N(x^{k,N}, \xi^{k,N}) := & \left(\frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} - \frac{\ell^{3/2}}{\sqrt{2}N^{3/4}} \right) \frac{x_j^{k,N} \xi_j^{k,N}}{\lambda_j} + \frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} \lambda_j \xi_j^{k,N} (\nabla \Psi^N(x^{k,N}))_j \\ & - \frac{\ell^2}{2N} (\xi_j^{k,N})^2 + I_2^N(x^{k,N}, y^{k,N}) + I_3^N(x^{k,N}, y^{k,N}) \,. \end{aligned} \tag{A.1}$$

We recall that I_2^N and I_3^N have been defined in Section 7. Therefore, using (7.8),

$$Q_{j,\perp}^N = Q^N - Q_j^N = I_1^N + \tilde{Q}_j^N, \tag{A.2}$$

having set

$$\tilde{Q}_j^N := -\left(\frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} - \frac{\ell^{3/2}}{\sqrt{2}N^{3/4}} \right) \frac{x_j^{k,N} \xi_j^{k,N}}{\lambda_j} - \frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} \lambda_j \xi_j^{k,N} (\nabla \Psi^N(x^{k,N}))_j + \frac{\ell^2}{2N} (\xi_j^{k,N})^2. \tag{A.3}$$

Proof of Lemma 7.6. (7.26) is a consequence of the definition (7.24) and the estimate (7.25). Thus, all we have to do is establish the latter. Recalling that $\{\hat{\phi}_j\}_{j\in\mathbb{N}} := \{j^{-s}\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}^s$, we act as in the proof of [PST12, Lemma 4.7] and obtain

$$\left| \left\langle \mathbb{E}_k \varepsilon^{k,N}, \hat{\phi}_j \right\rangle_s \right|^2 \lesssim j^{2s} \lambda_j^2 \mathbb{E}_k \left[Q_j^N(x^{k,N}, \xi^{k,N}) \right]^2$$

where Q_j^N has been defined in (A.1). Thus

$$
\begin{aligned}
\left|\left\langle \mathbb{E}_k \varepsilon^{k,N}, \hat{\phi}_j \right\rangle_s\right|^2 \lesssim & j^{2s}\lambda_j^2 \left(N^{-3/2} (x_j^{k,N})^2 \mathbb{E}_k \xi_j^2 \lambda_j^2 + N^{-5/2}\lambda_j^2 \mathbb{E}_k \left[\xi_j^2 (\nabla \Psi^N(x^{k,N}))_j^2 \right] \right) \\
& + j^{2s}\lambda_j^2 \mathbb{E}_k (\left|I_2^N\right|^2 + \left|I_3^N\right|^2) + \frac{j^{2s}\lambda_j^2}{N^2} \\
\lesssim & \, N^{-3/2} \mathbb{E}_k (j^s x_j^{k,N})^2 + N^{-5/2} j^{-2s} (\nabla \Psi^N(x^{k,N}))_j^2 \\
& + j^{2s}\lambda_j^2 N^{-2} + j^{2s}\lambda_j^2 \frac{1 + \left|\left|x^{k,N}\right|\right|_s^2}{\sqrt{N}},
\end{aligned}
$$

where the second inequality follows from the boundedness of the sequence $\{\lambda_j\}$, (7.13) and (7.14). Summing over j and applying (2.14) we obtain (7.25). $\square$

Proof of Lemma 7.7. By definition of $\varepsilon^{k,N}$, and because $\gamma^{k,N} = [\gamma^{k,N}]^2$ (as $\gamma^{k,N}$ can only take values 0 or 1)

$$
\mathbb{E}_k \left|\left|\varepsilon^{k,N}\right|\right|_s^2 = \sum_{j=1}^N j^{2s}\lambda_j^2 \mathbb{E}_k \left[\gamma^{k,N} \left| \xi_j^{k,N} \right|^2 \right] = \sum_{j=1}^N j^{2s}\lambda_j^2 \mathbb{E}_k \left[\left(1 \wedge e^{Q^N(x^{k,N}, y^{k,N})} \right) \left| \xi_j^{k,N} \right|^2 \right].
$$

Using the above, the Lipschitzianity of the function $s \mapsto 1 \wedge e^s$, (A.2) and the independence of $Q_{j,\perp}^N$ and $\xi_j^{k,N}$, we write

$$
\begin{aligned}
\left| \mathbb{E}_k \left|\left|\varepsilon^{k,N}\right|\right|_s^2 - \mathrm{Trace}(\mathcal{C}_s)\alpha_\ell(S^{k,N}) \right| &= \left| \mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left(1 \wedge e^{Q^N}\right) |\xi_j|^2 - \mathrm{Trace}(\mathcal{C}_s)\alpha_\ell(S^{k,N}) \right| \\
&\le \left| \mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left(1 \wedge e^{Q_{j,\perp}^N}\right) |\xi_j|^2 - \mathrm{Trace}(\mathcal{C}_s)\alpha_\ell(S^{k,N}) \right| \\
&+ \left| \mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left[\left(1 \wedge e^{Q^N}\right) - \left(1 \wedge e^{Q_{j,\perp}^N}\right) \right] |\xi_j|^2 \right| \\
&\lesssim \left| \sum_{j=1}^N j^{2s}\lambda_j^2 \mathbb{E}_k \left(1 \wedge e^{Q_{j,\perp}^N}\right) - \mathrm{Trace}(\mathcal{C}_s)\alpha_\ell(S^{k,N}) \right| \qquad \text{(A.4)} \\
&+ \left| \mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left|Q_j^N\right| |\xi_j|^2 \right| \qquad \text{(A.5)}
\end{aligned}
$$

We now proceed to bound the addends in (A.4) and (A.5), starting from the latter. Using (A.1) and (A.3), we write

$$
\begin{aligned}
\mathbb{E}_{x^0}\mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left|Q_j^N\right| |\xi_j|^2 \le & \, \mathbb{E}_{x^0}\mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left|I_2^N\right| |\xi_j|^2 + \mathbb{E}_{x^0}\mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left|I_3^N\right| |\xi_j|^2 \\
& + \mathbb{E}_{x^0}\mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 \left|\tilde{Q}_j^N\right| |\xi_j|^2 \\
\lesssim & \, \mathbb{E}_{x^0} \sum_{j=1}^N j^{2s}\lambda_j^2 (\mathbb{E}_k \left|I_2^N\right|^2)^{1/2} + \mathbb{E}_{x^0}\mathbb{E}_k \sum_{j=1}^N j^{2s}\lambda_j^2 (\mathbb{E}_k \left|I_3^N\right|^2)^{1/2} \\
& + \mathbb{E}_{x^0} \sum_{j=1}^N j^{2s}\lambda_j^2 \mathbb{E}_k \left(\left|\tilde{Q}_j^N\right| |\xi_j|^2 \right).
\end{aligned}
$$

The addends on the penultimate line of the above tend to zero thanks to Lemma 7.5, (2.7) and Lemma 7.2. As for the last addend, using (A.3):

$$\begin{aligned}\sum_{j=1}^{N} j^{2s}\lambda_j^2 \mathbb{E}_k\left[\left|\tilde{Q}_j^N\right| |\xi_j|^2\right] &\lesssim \frac{1}{N^{3/4}}\sum_{j=1}^{N} j^{2s}\lambda_j \left|x_j^{k,N}\right| \mathbb{E}_k\left|\xi_j^{k,N}\right|^3 \\ &\quad + \frac{1}{N^{5/4}}\sum_{j=1}^{N} j^{2s}\lambda_j^3 \left|(\mathcal{C}_N \nabla\Psi^N(x^{k,N}))_j\right| \mathbb{E}_k\left|\xi_j^{k,N}\right|^3 + \frac{1}{N}\sum_{j=1}^{N} j^{2s}\lambda_j^2 \mathbb{E}_k\left|\xi_j^{k,N}\right|^4 \\ &\lesssim \frac{1}{N^{3/4}}(1+\left|\left|x^{k,N}\right|\right|_s^2), \end{aligned} \tag{A.6}$$

where the last inequality follows from (2.15), (2.7), the boundedness of the sequence $\{\lambda_j\}_{j\in\mathbb{N}}$ and by using the Young Inequality (more precisely, the so-called Young inequlity "with ϵ"), as follows:

$$\lambda_j \left|x_j^{k,N}\right| \mathbb{E}_k\left|\xi_j^{k,N}\right|^3 \leq \left|x_j^{k,N}\right|^2 + \lambda_j^2\left(\mathbb{E}_k\left|\xi_j^{k,N}\right|^3\right)^2.$$

This concludes the analysis of the term (A.5). As for the term (A.4), by definition of α_ℓ, equation (1.11),

$$\begin{aligned}\left(1\wedge e^{Q_{j,\perp}^{k,N}}\right) - \alpha_\ell(S^{k,N}) &= \left(1\wedge e^{Q_{j,\perp}^{k,N}}\right) - \left(1\wedge e^{I_1^N(x^{k,N},y^{k,N})}\right) \\ &\quad + \left(1\wedge e^{I_1^N(x^{k,N},y^{k,N})}\right) - \left(1\wedge e^{\ell^2(S^{k,N}-1)/2}\right).\end{aligned}$$

Because $s\mapsto 1\wedge e^s$ is globally Lipschitz, using Lemma 7.5 and manipulations of the same type as in the above, we conclude that also (A.4) tends to zero as $N\to\infty$. This concludes the proof. $\square$

Appendix B: Uniform bounds on the moments of $S^{k,N}$ and $x^{k,N}$

Proof of Lemma 7.2. To prove both bounds, we use a strategy analogous to the one used in [PST14, Proof of Lemma 9]. Let $\{A_k : k\in\mathbb{N}\}$ be any sequence of real numbers. Suppose that there exists a constant $C\geq 0$ (independent of k) such that

$$A_{k+1} - A_k \leq \frac{C}{\sqrt{N}}\left(1+A_k\right). \tag{B.1}$$

We start by showing that if the above holds then $A_k \leq e^{CT}(A_0+CT)$, uniformly over $k=0,\dots,[T\sqrt{N}]$. Indeed, from (B.1),

$$A_k \leq \left(1+\frac{C}{\sqrt{N}}\right)^k A_0 + \frac{C}{\sqrt{N}}\sum_{j=0}^{k-1}\left(1+\frac{C}{\sqrt{N}}\right)^j \leq \left(1+\frac{C}{\sqrt{N}}\right)^k\left(A_0 + k\frac{C}{\sqrt{N}}\right).$$

Thus, for all $k=0,\dots,[T\sqrt{N}]$,

$$A_k \leq \left(1+\frac{C}{\sqrt{N}}\right)^{[T\sqrt{N}]}\left(A_0 + [T\sqrt{N}]\frac{C}{\sqrt{N}}\right) \leq \left(1+\frac{C}{\sqrt{N}}\right)^{T\sqrt{N}}(A_0+CT).$$

Since $[0,\infty)\ni N\mapsto (1+C/\sqrt{N})^{\sqrt{N}}$ is increasing,

$$\left(1+\frac{C}{\sqrt{N}}\right)^{\sqrt{N}} \leq \left(1+\frac{C}{\left\lceil\sqrt{N}\right\rceil}\right)^{\left\lceil\sqrt{N}\right\rceil} \leq \sum_{j=0}^{\left\lceil\sqrt{N}\right\rceil}\frac{C^j}{j!} \leq e^C.$$

With this preliminary observation, we can now prove (7.6) and

i) Proof of (7.6)**.** To prove (7.6) we only need to show that (B.1) holds (for some constant $C > 0$ independent of N and k) for the sequence $A_k = \mathbb{E}_{x^0}(S^{k,N})^q$. By the definition of $S^{k,N}$, we have

$$S^{k+1,N} = S^{k,N} + \frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^2_{\mathcal{C}_N}}{N} + \frac{2\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle_{\mathcal{C}_N}}{N}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_{x^0}(S^{k+1,N})^q - \mathbb{E}_{x^0}(S^{k,N})^q \\
&= \sum_{\substack{n+m+l=q \\ (n,m,l)\neq(q,0,0)}} \mathbb{E}_{x^0}\left[(S^{k,N})^n \left(\frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^2_{\mathcal{C}_N}}{N}\right)^m \left(\frac{2\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle_{\mathcal{C}_N}}{N}\right)^l\right]. \qquad \text{(B.2)}
\end{aligned}$$

Thus, to establish (B.1) it is enough to argue that each of the terms in the right-hand side of the above is bounded by $(C/\sqrt{N})(1 + \mathbb{E}(S^{k,N})^q)$. To this end, set

$$\begin{aligned}
J^{k,N} &:= \mathbb{E}_{x^0}\left[(S^{k,N})^n \left(\frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^2_{\mathcal{C}_N}}{N}\right)^m \left(\frac{2\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle_{\mathcal{C}_N}}{N}\right)^l\right] \\
&= \mathbb{E}_{x^0}\mathbb{E}_k\left[(S^{k,N})^n \left(\frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^2_{\mathcal{C}_N}}{N}\right)^m \left(\frac{2\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle_{\mathcal{C}_N}}{N}\right)^l\right].
\end{aligned}$$

By the Cauchy-Schwartz inequality for the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{C}_N}$,

$$\begin{aligned}
\frac{\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle^l_{\mathcal{C}_N}}{N^l} &\leq \frac{\left|\left|x^{k,N}\right|\right|^l_{\mathcal{C}_N}\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^l_{\mathcal{C}_N}}{N^l} \\
&= (S^{k,N})^{l/2}\frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^l_{\mathcal{C}_N}}{N^{l/2}},
\end{aligned}$$

which gives

$$J^N_k \lesssim (S^{k,N})^{n+l/2}\frac{\mathbb{E}_k\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^{2m+l}_{\mathcal{C}_N}}{N^{m+l/2}}.$$

Using the bound (7.4) of Lemma 7.1, we also have

$$\mathbb{E}_k\frac{\left|\left|x^{k+1,N} - x^{k,N}\right|\right|^{2m+l}_{\mathcal{C}_N}}{N^{m+l/2}} \lesssim \frac{(S^{k,N})^{m+l/2}}{N^{m+l/2}} + \frac{1}{N^{(m+l/2)/2}}.$$

Putting all of the above together (and using Young's inequality) we obtain

$$J^N_k \lesssim \frac{\mathbb{E}_{x^0}(S^{k,N})^q}{N^{m+l/2}} + \frac{1}{N^{m+l/2}}.$$

Now observe that $(m + l/2)/2 \geq 1/2$ except when $(n, m, l) = (q, 0, 0)$ or $(n, m, l) = (q - 1, 0, 1)$. Therefore we have shown the desired bound for all the terms in the expansion (B.2), except the one with $(n, m, l) = (q-1, 0, 1)$. To study the latter term, we recall that $\gamma^{k,N} \in \{0, 1\}$, and use the definition of the chain (equations (3.2) and (3.6)) to obtain

$$\begin{aligned}
\left|\left\langle x^{k+1,N} - x^{k,N}, x^{k,N}\right\rangle_{\mathcal{C}_N}\right| &\lesssim \delta\left|\left|x^{k,N}\right|\right|^2_{\mathcal{C}_N} + \delta\left|\left\langle \mathcal{C}_N \nabla\Psi^N(x^{k,N}), x^{k,N}\right\rangle_{\mathcal{C}_N}\right| \\
&\quad + \sqrt{\delta}\left|\left\langle x^{k,N}, (\mathcal{C}_N)^{1/2}\xi^{k,N}\right\rangle_{\mathcal{C}_N}\right|.
\end{aligned}$$

Combining (2.16) with the Cauchy-Schwartz inequality we have

$$\delta \left| \left\langle \mathcal{C}_N \nabla \Psi^N(x^{k,N}), x^{k,N} \right\rangle_{\mathcal{C}_N} \right| \lesssim N^{-1/2}(1 + \left|\left|x^{k,N}\right|\right|_s^2) \lesssim N^{-1/2} + N^{1/2} S^{k,N},$$

where in the last inequality we used the following observation

$$\left|\left|x^{k,N}\right|\right|_s^2 = \sum_{j=1}^{\infty} (x^{k,N})_j^2 j^{2s} = \sum_{j=1}^{\infty} \frac{(x^{k,N})_j^2}{\lambda_j^2} (\lambda_j^2 j^{2s}) \lesssim \sum_{j=1}^{\infty} \frac{(x^{k,N})_j^2}{\lambda_j^2} = N S^{k,N}.$$

Recalling that $\left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N}$, conditioned on $x^{k,N}$, is a linear combination of zero-mean Gaussian random variables, we have

$$\begin{aligned}
\mathbb{E}_k \sqrt{\delta} \left| \left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N} \right| &\lesssim 1 + N^{-1/2} \mathbb{E}_k \left| \left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N} \right|^2 \\
&\lesssim 1 + \sqrt{N} S^{k,N}.
\end{aligned}$$

Putting the above together and taking expectations we can then conclude

$$\begin{aligned}
\mathbb{E}\left[\frac{(S^{k,N})^{q-1} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_{\mathcal{C}_N}}{N} \right] &\lesssim \frac{\mathbb{E}\left[(S^{k,N})^{q-1}\right]}{N} + \frac{\mathbb{E}\left[(S^{k,N})^{q}\right]}{\sqrt{N}} \\
&\lesssim (1/\sqrt{N})(1 + \mathbb{E}\left[(S^{k,N})^{q}\right]),
\end{aligned}$$

and (7.6) follows.

ii) Proof of (7.7)**.** This is very similar to the proof of (7.6), so we only sketch it. Just as before, it is enough to establish the following bound

$$\mathbb{E}\left[\left|\left|x^{k,N}\right|\right|_s^{2n} \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^{2m} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l \right] \lesssim \frac{1}{\sqrt{N}} (1 + \mathbb{E}\left[\left|\left|x^{k,N}\right|\right|_s^{2q} \right])$$

for each (n, m, l) such that $n + m + l = q$ with the exception of the triple $(n, m, l) = (q, 0, 0)$. Applying the Cauchy-Schwartz inequality for $\langle \cdot, \cdot \rangle_s$ we have

$$\left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l \leq \left|\left|x^{k,N}\right|\right|_s^l \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^l.$$

Thus, Lemma 7.1 implies

$$\begin{aligned}
\mathbb{E}_k \left|\left|x^{k,N}\right|\right|_s^{2n} \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^{2m} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l &\leq \left|\left|x^{k,N}\right|\right|_s^{2n+l} \mathbb{E}_k \left|\left|x^{k+1,N} - x^{k,N}\right|\right|_s^{2m+l} \\
&\lesssim \frac{\left|\left|x^{k,N}\right|\right|_s^{2n+l} (1 + \left|\left|x^{k,N}\right|\right|_s^{2m+l})}{N^{(m+l/2)/2}}.
\end{aligned}$$

The above gives us the desired bound for all (n, m, l) except for $(n, m, l) = (q - 1, 0, 1)$. Like before, to study the latter case we observe

$$\begin{aligned}
\left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s &= \gamma^{k,N} (-\frac{1}{\sqrt{N}} (\left|\left|x^{k,N}\right|\right|_s^2 + \left\langle \mathcal{C}_N \nabla \Psi^N(x^{k,N}), x^{k,N} \right\rangle_s) \\
&\quad + \frac{\sqrt{2}}{N^{1/4}} \left\langle (\mathcal{C}_N)^{1/2} \xi^{k,N}, x^{k,N} \right\rangle_s) \\
&\lesssim \frac{1}{\sqrt{N}} (1 + \left|\left|x^{k,N}\right|\right|_s^2) + \frac{1}{N^{1/4}} \gamma^{k,N} \left\langle (C_N)^{1/2} \xi^{k,N}, x^{k,N} \right\rangle_s \\
&\lesssim \frac{1}{\sqrt{N}} (1 + \left|\left|x^{k,N}\right|\right|_s^2),
\end{aligned}$$

where penultimate inequality follows from the Cauchy-Schwartz inequality, (2.15), and the fact that $\gamma^{k,N} \in \{0, 1\}$, and the last inequality follows from Lemma 7.6. This concludes the proof.

□

Remark B.1. In [PST12] the authors derived the diffusion limit for the chain under weaker assumptions on the potential Ψ than those we use in this paper. Essentially, they assume that Ψ is quadratically bounded, while we assume that it is linearly bounded. If Ψ was quadratically bounded the proof of Lemma 7.6 would become considerably more involved. We observe explicitly that the statement of Lemma 7.6 is of paramount importance in order to establish the uniform bound on the moments of the chain x^k contained in Lemma 7.2. In [PST12] obtaining such bounds is not an issue, since the authors study the chain in its stationary regime. In other words, in [PST12] the law of $x^{k,N}$ is independent of k, and thus the uniform bounds on the moments of $x^{k,N}$ and $S^{k,N}$ are automatically true for target measures of the form considered there (see also the first bullet point of Remark 5.3). □

Acknowledgments M. Ottobre and J. Kuntz gratefully acknowledge financial support of the Edinburgh Mathematical Society. A.M. Stuart acknowledges support from AMS, DARPA, EPSRC, ONR.

References

[CRR05] O.F. Christensen, G.O. Roberts, and J.S. Rosenthal. Scaling limits for the transient phase of local Metropolis-Hastings algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(2):253–268, 2005.

[DZ92] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and Its Applications 44.* Number 1. Cambridge Univ. Press, Cambridge, 1992.

[HSV07] M. Hairer, A.M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling. Part II: the nonlinear case. *Ann. Appl. Probab.*, 17(5-6):1657–1706, 2007.

[HSVW05] M. Hairer, A.M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs Arising in path sampling. Part I: the Gaussian case. *Comm. Math. Sci.*, 3:587–603, 2005.

[JLM14] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of Metropolis Hastings algorithms: The longtime behavior. *Bernoulli*, 20(4):1930–1978, 2014.

[JLM15] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit. *The Annals of Applied Probability*, 25(4):2263–2300, 2015.

[KOS16] J. Kuntz, M. Ottobre, and A. M. Stuart. Diffusion limit for the Random Walk Metropolis algorithm out of stationarity. *Arxiv preprint*, 2016.

[MPS12] J.C. Mattingly, N.S. Pillai, and A.M. Stuart. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *The Annals of Applied Probability*, 22(3):881–930, 2012.

[OPPS16] M. Ottobre, N.S. Pillai, F.J. Pinski, and A.M. Stuart. A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli*, 22(1):60–106, 2016.

[PST12] N.S. Pillai, A.M. Stuart, and A.H. Thiéry. Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.*, 22(6):2320–2356, 2012.

[PST14] N.S. Pillai, A.M. Stuart, and A.H. Thiéry. Noisy gradient flow from a random walk in Hilbert space. *Stochastic Partial Differential Equations: Analysis and Computations*, 2(2):196–232, 2014.

[RGG97] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[RR98] G O Roberts and J S Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998.

[Stu10] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[Tie98] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.